



HAL
open science

Analyse de la méthylation de l'ADN par séquençage haut-débit chez la Poule

Marjorie Mersch

► **To cite this version:**

Marjorie Mersch. Analyse de la méthylation de l'ADN par séquençage haut-débit chez la Poule. Sciences agricoles. Institut National Polytechnique de Toulouse - INPT, 2018. Français. NNT : 2018INPT0107 . tel-02786411

HAL Id: tel-02786411

<https://hal.inrae.fr/tel-02786411>

Submitted on 29 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

Mme MARJORIE MERSCH

le mardi 30 octobre 2018

Titre :

Analyse de la méthylation de l'ADN par séquençage haut-débit chez la
Poule

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Génétique, Physiologie et Systèmes d'Élevage (GenPhySE)

Directeur(s) de Thèse :

MME FREDERIQUE PITEL

Rapporteurs :

Mme MARINA GOVOROUN, INRA TOURS

Mme PATRICIA THEBAULT, UNIVERSITÉ DE BORDEAUX

Membre(s) du jury :

M. ALAIN VIGNAL, INRA TOULOUSE, Président

M. JORDI ESTELLE-FABRELLAS, INRA JOUY EN JOSAS, Membre

Mme CATHERINE LABBE, INRA RENNES, Membre

Mme FREDERIQUE PITEL, INRA TOULOUSE, Membre

Table des matières

Remerciements	9
Index des tableaux	13
Index des figures	17
Glossaire	19
1 Introduction	23
1.1 Impact de l'environnement sur la variabilité des phénotypes	23
1.1.1 Impact des variations de température sur les productions avicoles	23
1.1.2 Compétition alimentaire	25
1.1.3 Variabilité phénotypique	27
1.1.4 Phénomènes épigénétiques	27
1.1.4.1 Modifications d'histones	27
1.1.4.2 Régulation par les ARN	29
1.1.4.3 Remaniement de la chromatine	31
1.2 Méthylation de l'ADN	31
1.2.1 Îlots CpG et régulation de l'expression génique	33
1.2.2 Méthylation de l'ADN pendant le développement	35
1.2.3 Héritabilité et transmission	37
1.2.4 Méthylation allèle-spécifique	37
1.2.5 Compensation de dose	39
1.2.6 Phénomènes épigénétiques chez les oiseaux : la revue	41
1.3 Données haut-débit et outils bioinformatiques	61
1.3.1 Amélioration de l'assemblage du génome de la poule grâce aux avancées technologiques	61
1.3.2 Données haut-débit du méthylome	63
1.3.2.1 Conversion au bisulfite	63
1.3.2.2 Séquençage haut-débit	63
1.3.3 Outils bioinformatiques d'analyses du méthylome	65
1.3.3.1 Pipeline d'analyses	65
1.3.3.2 Intégration de données -omiques	69
1.4 Objectifs de la thèse	71
1.4.1 Plan de la thèse	71
2 Dispositifs, données et développement bioinformatique	73
2.1 Méthylome d'embryons de poule	73
2.2 Réponse du méthylome à un stress environnemental	75
2.2.1 Protocole expérimental RRBS	75
2.3 Développement d'un pipeline bioinformatique	81
3 Analyses tout génome des données de méthylome	83
3.1 Projet de publication	83
3.2 Analyses complémentaires	109
3.2.1 Évolution des données	109
3.2.1.1 Évolution des versions du génome	109

3.2.1.2	Évolution du jeu de données	109
3.2.2	Contrôle de la qualité des données	109
3.2.2.1	Des lectures de séquençage aux CG "brutes"	109
3.2.2.2	Contrôle de la conversion au bisulfite	111
3.2.3	Choix des paramètres d'analyses	113
3.2.3.1	Définition d'un minimum de couverture	113
3.2.3.2	Méthodes d'annotation	113
3.2.3.3	Méthylation au TSS	115
3.2.3.4	Paramètres pour le différentiel de méthylation	119
3.2.4	Méthylation allèle-spécifique	119
3.2.4.1	ASM par croisement	125
3.2.4.2	ASM par haplotype	127
4	Analyses du méthylome en réponse à un stress environnemental	133
4.1	Contrôle de la qualité des données RRBS	133
4.1.1	Des lectures de séquençage aux CG analysés	133
4.1.2	Contrôle de la conversion au bisulfite	135
4.2	Méthylation des CG	137
4.2.1	Annotation des CG	137
4.2.2	Lien entre le taux de méthylation et le niveau d'expression	137
4.3	Différentiels de méthylation	143
4.3.1	Différentiel de méthylation entre lignées	143
4.3.2	Différentiel entre les contrôles et les stressés	145
4.3.2.1	Stress à la chaleur	145
4.3.2.2	Utilisation d'un aliment pauvre en énergie	149
4.3.2.3	Conclusion	151
5	Synthèse, conclusion générale, perspectives	153
5.1	Discussion sur les données et les méthodes utilisées	153
5.1.1	Données de séquençage	153
5.1.2	Pipeline d'analyses	159
5.2	Conclusion générale	161
5.3	Perspectives	165
A	Revue dans Productions Animales	171
	Bibliographie	191

Remerciements

Me voilà à la fin d'une aventure qui fut riche en émotions, en expériences, en rencontres, en fous rires, en découvertes... Je vous adresse à tous un très grand merci pour cette belle aventure!

Merci à l'Agence Nationale de la recherche (ANR) et au département de Génétique Animale de l'INRA pour avoir financé cette thèse.

Merci aux rapporteuses, Patricia Thébault et Marina Govoroun, ainsi qu'aux examinateurs, Catherine Labbé et Jordi Estellé, d'avoir évalué ce travail de thèse.

Merci aux membres de mon comité de thèse pour m'avoir prodigué des conseils et directions pour la thèse : Céline Noirot, Anis Djari, Alain Vignal et Vincent Coustham.

La thèse c'est comme gravir une montagne : on est plein de motivation et d'excitation lorsqu'on commence, mais sans vraiment se douter de ce qui nous attend, puis on arrive fatigué et heureux d'avoir réussi à gravir cette montagne! Mais ce qu'il faut souligner c'est que j'y suis arrivée en équipe avec les personnes qui m'ont encouragée, motivée, donnée des conseils, rassurée et formée, surtout par mes deux encadrants : Pitou et Sylvain. Je vous remercie chaleureusement pour m'avoir donné cette chance!

Pitou, merci vraiment, ta gentillesse, ton optimisme, et surtout ta passion pour la recherche m'ont montré que chercher, trouver, comprendre, se tromper et recommencer font parti du challenge, et surtout c'est amusant! On a eu des hauts et des bas pendant ces trois années et demi qui n'ont jamais entamé ton enthousiasme et ta passion de me faire apprendre la recherche. J'ai rencontré une belle personne, une chercheuse passionnée et une encadrante formidable.

Et une autre personne dans notre petite équipe que je voudrais remercier : Sylvain. Merci beaucoup également. Tu as animé cette thèse avec des idées toujours innovantes, pour ne pas dire parfois tirées par les cheveux, pour me faire comprendre, réfléchir et m'amener à m'améliorer. Tu as été disponible, patient et optimiste (et toujours prêt à faire des blagues) tout au long de cette thèse et tu m'as montré ce qu'un bioinfo apporte à une équipe de bio.

Je tenais vraiment à vous remercier tous les deux pour cette expérience inoubliable, cette thèse qui m'a beaucoup appris grâce à vous!

Sophie, sans toi toutes ces analyses n'auraient pas pu être possibles, merci. C'est indéniable que ce travail est un travail d'équipe dont tu as été un pilier central. Merci de ta disponibilité et d'avoir répondu à toutes mes questions de bioinfo pour comprendre le fonctionnement du labo.

Julie, merci beaucoup, pour ta gentillesse, tes fous rires et ton habitude de parler toute seule devant l'écran qui m'a bien fait rigoler! Partager un bureau avec une pas-si-ancienne-que-ça a été un super moment pour tout, y compris te poser des questions même quand tu étais submergée de boulot!

Annie, merci beaucoup, tu as toujours répondu présente pour m'aider et me donner des conseils. Malgré le sauna dans le bureau l'hiver (même si j'ai plutôt apprécié en fin de compte!), j'ai passé trois ans et demi dans ton bureau avec grand plaisir!

Et bien sûr un grand merci à tous les membres de mon équipe qui ont participé à ce travail de près ou de loin, mais toujours dans la bonne humeur : Yann, Katia, Maguy, Isabelle, Juliette, Guillaume, Nathalie, Amandine, Denis, Émilie et Noémie.

Je tenais également à adresser un grand merci à mon équipe barcelonaise qui m'a accueillie avec plaisir,

bonne humeur et disponibilité. Angelika, merci, tu m'as réservé un accueil chaleureux, et tu as toujours répondu à mes questions avec bonne humeur et sympathie, qu'elles soient sur la thèse ou d'ordre pratique. Et les restaurants étaient toujours au top!

Simon, merci de m'avoir accueillie au sein de ton équipe et montré un autre visage de la recherche, tout aussi passionnant.

Iris, merci à toi pour avoir partagé mes galères (deux étrangères dans une ville catalane!), discuté de tout et de rien, et avec qui j'ai partagé beaucoup de fous rires.

Juan, merci aussi de ta bonne humeur, de tes nombreuses discussions et ton enthousiasme tout espagnol.

Marcos, merci également, on a toujours réussi à communiquer malgré notre maîtrise parfois difficile de la langue anglaise.

Et je remercie également toutes les nombreuses personnes que j'ai côtoyé pendant mon bref mais ô combien sympathique passage à Barcelone.

Un grand merci également à l'équipe de GenoToul (oups genologin!) sans qui toutes les analyses bioinformatiques n'auraient pas pu se faire : Marie-Stéphane et Didier.

Également merci aux personnes de l'unité MIAT que j'ai pu rencontré, notamment Nathalie, Céline, Gaëlle, Alyssa.

Je souhaite aussi remercier mes partenaires athéniens, Malo, Lise, Gaëlle et Alyssa (toi aussi Sylvain!), pour avoir partagé mon premier congrès international dans une ville magnifique.

Merci aux gestionnaires de l'unité, qui sont toujours disponibles et répondent aux nombreuses questions des thésards avec bonne humeur : Florence, Nancy et Évelyne. Et merci à toi Manuëla!

J'aimerais terminer ce tour des remerciements du labo par les « jeunes » que j'ai rencontré. Un milliers de merci les filles (et Cyriel), cette thèse n'aurait pas été la même sans vous. D'abord mes randonneuses de l'extrême, Charlotte et Marjorie C, votre soutien, les folles soirées qu'on a faites, les randonnées, les voyages, vous connaître fait partie de mes plus belles découvertes toulousaines!

Noémie T, merci à toi, qui a toujours été prompte à rire, à soutenir, à aider avec générosité et surtout à me suivre dans mes escapades nocturnes! Tu es géniale!

Mathilde (Mathou!), merci à toi pour m'avoir montré le Toulouse que tu aimes, avec ces tapas, ces bières, ces soirées animées toutes aussi folles les unes que les autres.

Noémie G, merci beaucoup d'avoir apporté ta bonne humeur le temps d'un été, qui s'est ensuite concrétisé en amitié. Au prochain de nos voyages!

Morgane, merci à toi (et à Maël), pour ton enthousiasme, ta générosité, ton accueil et les fous rires qu'on a eu.

Pauline, merci de m'avoir accueilli chaleureusement, de m'avoir montré qu'une thésarde peut être au top.

Émilie, merci à toi, qui a vu beaucoup de la fin de ma thèse et peu du début. Tu verras l'expérience en vaut la peine!

Cyriel, merci, d'avoir partagé nos discussions de filles. Garde ton sens de l'humour!

Un énorme merci à mes super copines, Léa et Rania, on s'est croisé et on ne s'est plus quitté!

Cette belle aventure n'aurait pas été la même sans le soutien de ma famille. Merci à mes parents. À ma sœur, Victoria, toi et moi c'est pour la vie! À mon petit grand frère, Henri, tu le sais on aura toujours le même cerveau.

À Loïc qui me soutient depuis peu mais qui est déjà tellement présent dans ma vie.

Index des tableaux

1.1	Principales modifications post-traductionnelles d’histones.	28
1.2	Informations sur l’assemblage des versions successives de <i>Gallus gallus</i>	60
2.1	Dispositif du projet EpiBird (données WGBS)	74
2.2	Dispositif du projet ChickStress (données RRBS)	74
2.3	Fenêtres de taille des fragments	78
2.5	Échantillons d’hypothalamus d’individus adultes issus des données RRBS	79
2.6	Ressources informatiques utilisées par le pipeline	82
3.1	Nombre de lectures sur les chromosomes sexuels	108
3.2	Pourcentage de lectures après chaque étape du pipeline	111
3.3	Taux de conversion des C hors CG	112
3.4	Taux de méthylation moyen du plasmide	113
3.5	Nombre de DMC et de DMR pour chaque condition de différentiel	118
3.6	Données alignées sur le double génome	123
3.7	Nombre de SNP des génomes parentaux	123
3.8	Principales fonctions des gènes associés aux DMC par haplotype.	128
4.1	Données issues de l’hypothalamus	133
4.2	Contrôles de la conversion au bisulfite	136
4.3	Nombre de DMC et de DMR pour le différentiel entre lignées	144
4.4	Nombre de DMC et de DMR pour le différentiel stressé à la chaleur	146
4.5	Nombre de DMC et de DMR pour le différentiel carencé en énergie	148
5.1	Comparaison des données WGBS de poule	160

Index des figures

1.1	Modification chimique des ARN messagers	30
1.2	Conversion d'une cytosine en 5-méthylcytosine	30
1.3	Dérivés de l'oxydation de la 5-méthylcytosine	30
1.4	Schéma des îlots CpG et de la régulation de l'expression associée	34
1.5	Schéma du phénomène de méthylation allèle-spécifique	36
1.6	Schéma du principe de la conversion au bisulfite de sodium	62
1.7	Schéma de l'alignement d'une lecture traitée au bisulfite	64
1.8	Visualisation des lectures alignées sur la référence	66
1.9	Analyses intégratives de la variabilité phénotypique selon plusieurs types de données	68
2.1	Schéma du dispositif expérimental du projet EpiBird	74
2.2	Étapes du séquençage RRBS	76
2.3	Distribution des CG dans les fragments digérés par MspI	78
2.4	Distribution des intervalles de taille des fragments après digestion enzymatique	78
2.5	Schéma des enchaînements de scripts du pipeline	80
3.1	Visualisation des lectures alignées sur le chromosome W	108
3.2	PCA des huit embryons sur les CG détectés	110
3.3	Distribution des lectures	110
3.4	Taux de méthylation des C du plasmide en WGBS	112
3.5	Distribution de la couverture des CG dans chaque embryon	114
3.6	Schéma de deux CG dans les caractéristiques géniques	114
3.7	Schéma d'un îlot CpG, des <i>shores</i> , des <i>shelves</i> et de l' <i>openSea</i>	116
3.8	Schéma d'une même position de CG dans plusieurs régions génomiques	116
3.9	Nombre de CG dans les deux méthodes d'annotation	116
3.10	Taux de la méthylation autour du TSS	117
3.11	Schéma de la fonction de lissage en différentiel de méthylation	118
3.12	Nombre de DMC et de DMR selon les fenêtres de lissage	118
3.13	Schéma de la méthylation allèle-spécifique détectée sur un double génome.	120
3.14	Distribution du MAPQ sur le génome de référence et sur le double génome	120
3.15	Distribution du MAPQ des mêmes lectures alignées sur le génome de référence et sur le double génome	122
3.16	Méthylation différentielle des DMC communes aux deux sens de croisement	124
3.17	Distance entre les SNP et les DMC	124
3.18	Méthylation par allèle selon les sens de croisements	124
3.19	Méthylation et de l'expression différentielles au sein de chaque croisement.	126
3.20	Méthylation par allèle	126
3.21	Distribution de la méthylation et de l'expression différentielles par haplotype	129
3.22	Méthylation ASM et expression ASE	130
4.1	Distribution des CG dans les caractéristiques génomiques	138
4.2	Distribution de la méthylation des CG dans les caractéristiques génomiques	138
4.3	Répartition des CG et leur taux de méthylation dans les caractéristiques géniques	139
4.4	Répartition des CG et leur taux de méthylation dans les îlots CpG	139
4.5	Taux de méthylation moyen dans les îlots CpG selon les caractéristiques génomiques pour tous les individus	140

4.6	Taux de méthylation autour du TSS sur tous les individus RRBS	141
4.7	Taux de méthylation selon les quantiles d'expression en RRBS	141
4.8	Distribution du taux de méthylation selon les niveaux d'expression dans le gène	142
4.9	Catégories biologiques des DMC entre R^+ et R^- liés à des gènes	144
4.10	Distribution de la méthylation des CG et des DMC pour les conditions CT et HS	146
4.11	Répartition des DMC issues du différentiel CT/HS	147
4.12	Enrichissement des DMC dans le gène pour le différentiel CT/HS	147
4.13	Distribution de la méthylation des CG et des DMC pour les conditions CT et FS	148
4.14	Répartition des DMC issues du différentiel CT/FS	148
5.1	Comparaison des positions de CG filtrés issues de RRBS et WGBS	154
5.2	Comparaison des positions détectées et annotées	156
5.3	Distribution des CG issus des deux techniques de séquençage	156
5.4	Distribution de la méthylation des CG communs RRBS/WGBS à partir des CG WGBS	157
5.5	Visualisation d'une région génomique analysée en RRBS et WGBS.	157
5.6	Redondance de l'information de méthylation sur des lectures <i>paired-end</i>	158
5.7	Étapes du pipeline gemBS	158
5.8	Étapes du pipeline gemBS adapté aux données WGBS de poule	160
5.9	Différence du taux de méthylation pour un individu entre les deux pipelines	162

Glossaire

3'UTR : UnTranslated Region en 3'

5'UTR : UnTranslated Region en 5'

ANR : Agence Nationale de la Recherche

BCAS3 : Breast Carcinoma Amplified Sequence 3

BRMS1L : Breast Cancer Metastasis-Suppressor 1 Like

BS-seq : Bisulfite Sequencing

CD93 : Human cluster of Differentiation 93

CDS : Coding-Sequence

ChIP-seq : Chromatin ImmunoPrecipitation sequencing

CNV : Copy Number Variation

CpG : Cytosine-phosphate-Guanine

CTOB : Complementary To Original Bottom

CTOT : Complementary To Original Top

DMC : Differentially Methylated Cytosine

DMR : Differentially Methylated Region

DNAJC16 : DNAJ Heat Shock Protein Family (Hsp40) Member C16

DNA-seq : DNA sequencing

DNMT : DNA Methyltransferases

FAO : Food and Agriculture Organization of the United Nations

FS : Feed Stress

Go : Giga-octet

GWAS : Genome-Wide Association Studies

HHS : Axe Hypothalamique Hypophysaire Surrénalien

HS : Heat Stress

IPCC : Intergovernmental Panel on Climate Change

ISPD : Isoprenoid Synthase Domain Containing

ITAVI : Institut Technique de l'Aviculture

JKAMP : JNK1/MAPK8 Associated Membrane Protein

MHM : Male HyperMethylated

MRC2 : Mannose Receptor C Type 2

NGS : Next-Generation Sequencing

NOMe-seq : Nucleosome Occupancy and Methylome sequencing
OB : Original Bottom
OT : Original Top
PacBio : Pacific Biosciences RSII
pb : paire de bases
PCR : Polymerase Chain Reaction
PRB : Population Reference Bureau
PRRX1 : Paired Related Homeobox 1
PTM : Post-Translational Modifications
QTL : Quantitative Trait Loci
RJF : Red Jungle Fowl
RNA-seq : RNA sequencing
RRBS : Reduced Representation Bisulfite Sequencing
SAM : Sequence Alignment/Map
SGE : Sun Grid Engine
SLURM : Simple Linux Utility for Resource Management
SNP : Single Nucleotide Polymorphism
TDG : Thymine-DNA Glycosylase
TET1 : Ten-Eleven Translocation methylcytosine dioxygenase 1
TSS : Transcription Start Site
TTS : Transcription Termination Site
UDG : Uracil-DNA Glycosylase
WGBS : Whole Genome Bisulfite Sequencing

Chapitre 1

Introduction

1.1 Impact de l'environnement sur la variabilité des phénotypes

La question du changement climatique est aujourd'hui un enjeu sociétal majeur. Ce phénomène a été observé dès le début du XX^e siècle. L'évolution des températures moyennes mondiales (+0,85°C entre 1880 et 2012), le réchauffement des océans (+0,71°C entre 1971 et 2010), la diminution de moitié de la cryosphère (neige et glace) et l'augmentation du niveau des mers (+19 cm entre 1901 et 2010) ont conduit à la conclusion d'un phénomène de réchauffement planétaire selon le cinquième rapport de l'IPCC/GIEC (*Intergovernmental Panel on Climate Change*; Groupe d'experts Intergouvernemental sur l'Évolution du Climat) [1]. Ce rapport a démontré que ce phénomène est "très probablement" dû aux activités humaines intenses de ces deux derniers siècles. L'actuelle prédiction de l'évolution du climat a montré un scénario pessimiste mais probable correspondant à la prolongation des émissions actuelles de CO₂. En plus des secteurs énergétique, industriel et forestier, l'agriculture a vu son activité augmenter considérablement jusqu'à produire 13,5% des émissions de CO₂ de la planète selon le quatrième rapport de l'IPCC [2]. Cette agriculture intensive utilise de plus en plus de machines (ou robots) et d'intrants agricoles ce qui a de forts impacts environnementaux. Paradoxalement, ces changements climatiques ont des conséquences directes sur l'agriculture comme une diminution du rendement des cultures au niveau mondial (maïs -1,2%, blé -2% [1]) due à une chaleur accrue parfois extrême, des pénuries alimentaires, l'augmentation de ravageurs et des maladies et des mauvaises récoltes provoquées par la baisse des précipitations. Un des grands défis pour demain est de développer de nouvelles approches agronomiques adaptées aux problématiques environnementales, notamment dans le domaine des productions animales où l'élevage est responsable de 14,5% des émissions mondiales de gaz à effet de serre selon la FAO (*Food and Agriculture Organization of the United Nations* [3]).

1.1.1 Impact des variations de température sur les productions avicoles

La production avicole est une source majeure de protéines dans l'alimentation humaine dont la population ne cesse de croître (+1,2% par an selon le PRB (*Population Reference Bureau*)). Selon la FAO (<http://faostat.fao.org/>), après le porc, la volaille est la deuxième viande la plus consommée au monde avec 110,5 millions de tonnes en 2014 et la production d'œufs de poules dans le monde atteint 69,8 millions de tonnes en 2014. La France est le premier producteur d'œufs en Europe (14,3 milliards en 2016). Lors des saisons avec des fortes chaleurs, les animaux ont des difficultés à maintenir une température corporelle constante et une production constante sans efforts, quand ils survivent. L'épisode caniculaire de l'été 2003 en France, par exemple, a été à l'origine de 5,7 millions de morts dans les élevages de volailles, soit une perte d'environ 44,5 millions d'euros [4]. Ainsi, il devient crucial de contrôler les conditions environnementales, et/ou de sélectionner des individus qui s'adaptent mieux pour maintenir la production et le bien-être des animaux. Les poules pondeuses sont particulièrement sensibles aux pics de chaleur en raison de leur cycle de production long de 52 à 54 semaines (<https://www.itavi.asso.fr>). L'exposition à de fortes températures entraîne une baisse de la consommation alimentaire avec une perte de productivité et l'augmentation du taux de mortalité [5][6]. Cette baisse de consommation entraîne une baisse du calibre des œufs, de leur nombre et de la qualité de la coquille (épaisseur et solidité).

La température interne des poules se situe entre 41°C et 42°C et leur zone de confort thermique est

entre 18°C et 24°C [7]. La sensibilité des poules pondeuses à la chaleur provient du fait que leur production métabolique de chaleur est élevée, la dissipation thermique par la convection et la radiation est faible et leur taux d'évaporation d'eau par la respiration n'est pas suffisant pour maintenir l'homéothermie en raison de l'absence de glandes sudoripares [8]. À cause de cette absence physiologique, les poules ont une faible capacité à perdre de la chaleur par évaporation via la surface de la peau. Pour maintenir la thermorégulation, des sacs aériens (organes du système respiratoire) et une grande surface d'évaporation des poumons permettent de refroidir le corps par halètement [9]. De plus, la présence de plumes réduit fortement la sudation. La variabilité génétique des gènes ayant un rôle dans la présence de plumes est associée à une tolérance à la chaleur plus ou moins importante. Par exemple, les poules portant le gène "cou nu" (*Naked Neck*) qui limite l'extension du plumage supportent mieux la chaleur [5]. Certains paramètres sanguins sont également de bons indicateurs du stress à la chaleur comme l'augmentation du niveau de CO₂ et du pH sanguin (alcalose) ce qui empêche la disponibilité du bicarbonate pour la minéralisation de la coquille et diminue le taux de calcium libre dans le sang. Cela affecte la qualité de la coquille [10]. De plus, le stress à la chaleur affecte les fonctions des hormones reproductives dans l'hypothalamus et dans l'ovaire chez les femelles. Chez les mâles, le volume du liquide séminal, la concentration en sperme, le nombre de spermatozoïdes vivants et leur motilité sont affectés [9]. Les concentrations de corticostérone du plasma, un indicateur de l'axe hypothalamique hypophysaire surrénalien (HHS, appelé aussi axe du stress) sont augmentées [11] tandis que la triiodothyronine (T₃), hormone thyroïdienne du métabolisme et impliquée dans la thermorégulation, diminue lors de fortes températures [12]. Ainsi, en diminuant l'ingestion alimentaire et en activant l'axe HHS (modification du profil neuroendocrinien), la production d'œufs et de viandes de volailles est altérée.

1.1.2 Compétition alimentaire

En plus de cette baisse de production due aux variations de températures, une compétition alimentaire peut s'installer sur des aliments communs à l'alimentation des poules et à celle des Hommes [13]. Le prix des cultures exploitées pour la production animale devrait monter avec la diminution du rendement. Ainsi une nouvelle répartition des protéines consommables par l'Homme et l'animal se dessine. Cependant, les animaux d'élevage mangent également des coproduits non consommables par l'Homme. Par exemple, 45% de protéines utilisées dans la production de poulets de chair "standard" peuvent être consommables par l'Homme. Néanmoins, la production restitue 63% de protéines consommables par l'Homme (sur 100 g de poulet, 63 g sont consommés par l'Homme). En poule pondeuse, 28% de l'alimentation des animaux utilisée en production pourraient être consommées par l'Homme. Cependant, nous consommons 97,5 œufs sur 100 issus de la production et 53% de protéines sont restituées sous forme de poules de réforme [13]. Valoriser les coproduits végétaux non consommables par l'Homme peut donc contribuer à une augmentation de protéines disponibles pour l'alimentation humaine. En outre, l'alimentation représente environ 60% à 65% du coût d'une production avicole selon l'ITAVI (Institut Technique de l'AViculture), utiliser plus de coproduits moins coûteux aurait donc un intérêt économique majeur.

Un régime suffisant en teneur protéique et énergétique est nécessaire aux poules pondeuses adultes pour maintenir leur croissance (poids de l'animal) et maximiser la production d'œufs (poids et qualité) [14][15]. Un régime basse énergie entraîne une augmentation de la consommation alimentaire qui peut se révéler insuffisante pour obtenir le même apport énergétique qu'un régime à haute teneur énergétique. Lorsque du matériel à haute teneur énergétique est remplacé par une alimentation moins énergétique, le poids de l'œuf diminue [16][17][18]. La compensation de cet apport moindre d'énergie peut se faire par une augmentation de la consommation alimentaire, ou une sélection de lignées à plus faible consommation alimentaire résiduelle. La consommation alimentaire résiduelle est la différence entre la consommation alimentaire observée et la consommation alimentaire théorique qui est calculée en fonction des besoins de croissance et d'entretien de l'organisme [19][20]. En plus du niveau énergétique du régime, sa composition a un impact sur les performances de l'animal. Lorsqu'on modifie le niveau énergétique du régime dans le but d'utiliser des coproduits par exemple, il est important de garder la même composition en lipides et acides gras. Les changements dans la composition en acides gras affectent la proportion entre l'albumen et le jaune. Une augmentation du taux d'acides gras dans le jaune a, de plus, un impact bénéfique sur la santé humaine [21]. De plus, les fibres sont souvent utilisées en complément dans les régimes à basse teneur protéique afin de réduire les aliments riches en protéines. Il a été montré que l'ingestion d'une quantité modérée de fibres peut améliorer la croissance et la santé des poules [22]. Ainsi, un nombre croissant d'études s'intéressent à l'utilisation de coproduits dans l'alimentation des poules pour conserver des performances de production sans rentrer en compétition avec l'Homme sur les mêmes ressources

alimentaires.

1.1.3 Variabilité phénotypique

La variabilité phénotypique est gouvernée par la variabilité génétique, l'environnement et leur interaction. La variabilité génétique d'une population représente les différents polymorphismes génétiques présents dans la population. Elle est utilisée pour sélectionner des caractères d'intérêt en production animale comme par exemple une meilleure robustesse [23]. Un caractère peut être sous déterminisme génétique total : étudier les phénotypes permet "l'accès direct" aux variants génétiques. Ces variants entrent dans la catégorie des "gènes majeurs", pour lesquels la variabilité phénotypique est en grande partie expliquée par un seul locus. Pour la plupart des phénotypes d'intérêt, ou caractères complexes, la variabilité phénotypique est gouvernée par un grand nombre de loci, appelés QTL (*Quantitative Trait Loci*) : plusieurs régions du génome sont à l'origine de la variabilité du caractère quantitatif, chacune ayant un effet minime. La variabilité génétique peut reposer sur des mutations soit de petite taille (un à quelques nucléotides), ce sont les variants génétiques, comme les SNP (*Single Nucleotide Polymorphism*) et les petits indels (insertion / délétion), soit de plus grande taille (des dizaines à des millions de nucléotides), ce sont des variants structuraux, comme les indels, les duplications et les inversions de segments chromosomiques. Découvrir les mutations causales responsables de la variabilité des caractères d'intérêt peut permettre d'améliorer les schémas de sélection dans les populations d'animaux d'élevage. De nombreux QTL ont été identifiés comme associés à des caractères agronomiques d'intérêt [24]. L'association de la variabilité d'un marqueur moléculaire avec un phénotype d'intérêt en des positions du génome, mise en évidence par recherche de QTL ou GWAS (*Genome-Wide Association Studies*), relie l'information génétique et le phénotype. En plus de la variabilité génétique, l'environnement joue un rôle important dans la variabilité phénotypique : la plasticité phénotypique d'un individu repose en partie sur des facteurs non-génétiques qui ont un impact sur le phénotype sans qu'il y ait forcément une transmission de cette adaptation aux générations suivantes. Une partie de la réponse du génome aux changements environnementaux est liée à des phénomènes épigénétiques. Ainsi l'un des objectifs actuels des recherches en productions animales est d'estimer quelle est la part des causes épigénétiques dans la variabilité phénotypique observée chez les espèces d'élevage.

1.1.4 Phénomènes épigénétiques

Les phénomènes épigénétiques peuvent intervenir dans la régulation de l'expression des gènes et sont en partie sensibles à l'environnement [25]. Ce sont des mécanismes qui n'affectent pas la séquence d'ADN mais qui peuvent être transmis par la mitose ou la méiose. Le paysage épigénétique se compose de plusieurs mécanismes moléculaires, tels que la méthylation de l'ADN [26], les modifications post-traductionnelles des histones [27], la régulation de l'expression des gènes par les ARN [28], le remaniement de la chromatine et la stabilité du génome (cf revues, voir [29][30][31][32]).

La méthylation de l'ADN en tant que marque épigénétique est au cœur de ce travail. Elle sera détaillée dans le paragraphe suivant (cf partie 1.2 page 31). Les modifications d'histones, la régulation épigénétique par les ARN et le remaniement de la chromatine seront ici décrits succinctement, la description de chacun de ces phénomènes ainsi que leurs interactions étant hors du propos de cette thèse. Des revues récentes sont référencées pour chaque paragraphe.

1.1.4.1 Modifications d'histones

Un des mécanismes épigénétiques jouant sur la régulation de l'expression des gènes est la modification post-traductionnelle des histones (PTM, *Post-Translational Modification(s)*). Chez les eucaryotes, dans le noyau cellulaire, l'ADN est compacté et ramassé en un complexe moléculaire appelé chromatine dont l'unité fondamentale est le nucléosome, qui est associé avec une protéine de liaison appelée histone H1. Le nucléosome est composé d'un octamère de protéines d'histones (2 unités de H2A, H2B, H3, H4). Les extrémités N-terminales des histones peuvent être ciblées par des enzymes qui vont induire des modifications post-traductionnelles (acétylation, méthylation, phosphorylation et ubiquitinylation) influençant la conformation de la chromatine, le recrutement ou l'éviction des protéines et la transcription [33]. Ces modifications ont également un rôle important dans les processus de l'ADN comme la réparation, la réplication et la recombinaison. Les voies de régulation des histones sont associées, de façon dépendante en partie, à la méthylation de l'ADN. En effet, leur renforcement mutuel implique des interactions entre des enzymes et des facteurs associés [27]. Les

Tableau 1.1 – Principales modifications post-traductionnelles d’histones. Ces modifications d’histones sont impliquées dans différentes fonctions biologiques chez les vertébrés. Les modifications d’histones présentes uniquement chez les mammifères ne sont pas notées ici. Dernière mise à jour en mars 2018 (Tableau issu de [37] et [38]).

Histone	Site	Modification	Enzymes	Fonctions	
H1	Lys26	Méthylation	Ezh2	Répression de la transcription	
	Ser27	Phosphorylation	Non connues	Activation de la transcription, décondensation chromatinienne	
H2A	Arg3	Méthylation	PRMT1/6, PRMT5/7	Activation de la transcription, répression de la transcription	
	Ser1	Phosphorylation	MSK1	Mitose, assemblage de la chromatine, répression de la transcription	
	Ser139	Phosphorylation	ATR, ATM, DNA-PK	Réparation de l'ADN	
	Thr142	Phosphorylation	WSTF	Apoptose, réparation de l'ADN	
	Lys13	Biotinylation	biotinidase	Non connues	
H2B	Lys5	Acétylation	p300, ATF2	Activation de la transcription	
	Lys20	Acétylation	p300	Activation de la transcription	
	Ser14	Phosphorylation	Mst1	Apoptose, réparation de l'ADN	
	Ser36	Phosphorylation	AMPK	Activation de la transcription	
	Lys120	Ubiquitylation	UbcH6	Méiose	
H3	Lys9	Acétylation	Gcn5, SRC-1	Déposition des histones, activation de la transcription	
	Lys14	Acétylation	Gcn5, PCAF, Esa1, Tip60, SRC-1, Eip3, Hpa2, hTFIIIC90, TAF1, Sas2, Sas3, p300	Déposition des histones, activation de la transcription, réparation de l'ADN, transcription ARN polymérase III, transcription ARN polymérase II, euchromatine	
	Lys18	Acétylation	Gcn5, p300/CBP	Activation de la transcription, réparation de l'ADN, réplication de l'ADN	
	Lys23	Acétylation	Gcn5, Sas3, p300/CBP	Déposition des histones, activation de la transcription, réparation de l'ADN	
	Lys27	Acétylation	p300/CBP	Activation de la transcription	
	Lys36	Acétylation	Gcn5	Activation de la transcription	
	Arg2	Méthylation	PRMT5, PRMT6	Répression de la transcription	
	Arg8	Méthylation	PRMT5, PRMT2/6	Activation de la transcription, répression de la transcription	
	Arg17	Méthylation	CARM1	Activation de la transcription	
	Arg26	Méthylation	CARM1	Activation de la transcription	
	Arg42	Méthylation	CARM1	Activation de la transcription	
	Lys4	Méthylation	Set 7/9, MLL, ALL-1	Activation de la transcription	
	Lys9	Méthylation	Suv39h, Clr4, G9a, SETDB1	Répression de la transcription (tri-Me), répression transcriptionnelle de l'empreinte parentale génomique	
	Lys27	Méthylation	Ezh2, G9a	Répression de la transcription, inactivation de l'X (tri-Me)	
	Lys36	Méthylation	Set2	Activation de la transcription (élongation)	
	Lys79	Méthylation	Dot1	Euchromatine, activation de la transcription (élongation)	
	Ser10	Phosphorylation	Aurora-B kinase, MSK1, MSK2, IKK- α , Snf1	Mitose, méiose, activation génique immédiate-précoce, activation de la transcription	
	Thr3	Phosphorylation	Haspin/Gsg2	Mitose	
	Thr6	Phosphorylation	PKC β I	Déméthylation de H3K4	
	Tyr41	Phosphorylation	JAK2	Activation de la transcription	
	Tyr45	Phosphorylation	PKC δ	Apoptose	
	Lys4	Biotinylation	biotinidase	Expression des gènes	
	Lys9	Biotinylation	biotinidase	Expression des gènes	
	Lys18	Biotinylation	biotinidase	Expression des gènes	
	H4	Lys5	Acétylation	Hat1, Esa1, Tip60, ATF2, Hpa2, p300	Déposition des histones, activation de la transcription, réparation de l'ADN
		Lys8	Acétylation	Gcn5, PCAF, Esa1, Tip60, ATF2, Eip3, p300	Activation de la transcription, réparation de l'ADN
Lys12		Acétylation	Hat1, Esa1, Tip60, Hpa2, p300	Déposition des histones, répression télomérique, activation de la transcription, réparation de l'ADN	
Lys16		Acétylation	Gcn15, Esa1, Tip60, ATF2, Sas2	Activation de la transcription, réparation de l'ADN, euchromatine	
Arg3		Méthylation	PRMT1/6, PRMT5/7	Activation de la transcription, répression de la transcription	
Lys20		Méthylation	PR-Set7, Suv4-20h	Répression de la transcription (mono-Me), hétérochromatine (tri-Me)	
Lys59		Méthylation	Non connu	Répression de la transcription	
Ser1		Phosphorylation	CK2	Mitose, assemblage de la chromatine, réparation de l'ADN	
Lys12		Biotinylation	biotinidase	Réponse aux dommages de l'ADN	

modifications post-traductionnelles peuvent affecter directement la structure de la chromatine ou agir sur des sites de fixation pour les protéines ayant des domaines de liaison spécifiques. Des études ont montré que l'environnement influence les modifications des histones (cf revue [34]). Par exemple, dans le cerveau de souris exposées avant la naissance à l'arsenic, il a été observé qu'en plus de l'hypo-acétylation globale de H3K9 (histone 3 lysine 9), les données ChIP-seq suggéraient une représentation enrichie des régions promotrices de certains facteurs de transcription en doigts de zinc [35]. Un autre exemple est le phénomène de vernalisation chez les plantes [36]. Chez *Arabidopsis thaliana*, le rôle de H3K27me3 dans le contrôle épigénétique du phénotype est illustré par la répression des gènes induite par le froid. Cet exemple a été décrit dans la revue intégrée à ce travail de thèse (cf partie 1.2.6 page 41), qui décrit également plus en détails les marques d'histones et les méthodes de leur étude chez le poulet. Un tableau récapitulatif des principales marques d'histones et leurs fonctions étudiées est représenté (Tableau 1.1).

1.1.4.2 Régulation par les ARN

Les ARN longs non codants (longs ncRNA) régulent les processus biologiques de l'ADN comme la transcription (cf revues, voir [39][40]). Leur rôle est de contrôler l'expression des gènes au niveau transcriptionnel et post-transcriptionnel par des mécanismes épigénétiques [41], et notamment durant le développement. En outre, des études ont montré le rôle des ncRNA dans la réponse à un stress environnemental (cf revues, voir [42] pour les plantes et [43] pour tous les organismes). Bien que les longs ncRNA soient stables, généralement ils le sont moins que les ARN messagers qui codent pour des protéines [44]. L'utilité de la régulation de la biosynthèse et de la stabilité des ncRNA a été clairement montrée dans les processus qui requièrent des changements dynamiques ou rapides chez les différentes espèces, comme l'adaptation à la lumière de la rétine [45] ou les changements de rythme circadien [46] [47] [48]. De plus, certains ncRNA ne sont nécessaires que de façon transitoire, comme par exemple p15AS (*p15 antisense RNA*) qui initie la répression du gène p15 par une inhibition de la chromatine, mais n'est pas requis pour la maintenance de la répression [49]. Cet exemple souligne l'interaction entre la régulation des ARN avec les mécanismes épigénétiques comme un moyen d'établir des effets de régulation stables. Beaucoup d'exemples de ncRNA impliqués dans la réponse au stress ont été décrits chez les animaux [50]. Chez la drosophile, le locus de choc thermique *hsr- ω* (*Heat Shock RNA omega*) produit un ensemble de longs ncRNA localisés dans le noyau. Les ARN *hsr- ω* sont exprimés dans plusieurs tissus, y compris pendant le développement, et forment des complexes nucléaires localisés appelés tâches ω , qui sont répartis dans l'espace inter-chromatinien à proximité de la chromatine [51][52][53][54]. Ces ncRNA ont un rôle dans la régulation de la disponibilité des ribonucléoprotéines hétérogènes nucléaires des compartiments actifs (liés à la chromatine) et inactifs (dans les tâches ω) et interfèrent avec les pré ARN messagers [55]. Il est également intéressant de noter que *hsr- ω* est aussi induit par d'autres facteurs de stress, tels que les amides, les benzamides et la colchicine [56], ce qui souligne les nombreux rôles des ncRNA dans la réponse aux stress. Chez les mammifères, plusieurs transcrits non codants ont été également impliqués dans la réponse aux stress, comme les ARN de choc thermique, et ont été montrés comme orchestrant des changements globaux à des niveaux transcriptionnels et post-transcriptionnels. La transcription par l'ARN polymérase III d'éléments répétés SINE (*Short interspersed nuclear elements*) chez la souris et chez l'Homme augmente considérablement pendant un stress thermique et favorise une répression transcriptionnelle générale par une inhibition de l'ARN polymérase II, en interférant avec la formation du complexe de pré-initiation sur les promoteurs de gènes spécifiques [57][58]. De plus, les éléments répétés péricentriques satellites III sont activés par un stress thermique ou chimique, générant une grande quantité de ncRNA poly-adiénylés qui s'accumulent à l'intérieur du noyau [59][60]. Un autre ncRNA, HSR1 (*Heat Shock RNA-1*), a été identifié comme un ARN thermosensible dans les cellules des mammifères. Sous l'effet d'un choc thermique, HSR1 semble subir un changement conformationnel, puis forme un complexe avec le facteur d'élongation de traduction eEF1A (*Eukaryotic translation elongation factor 1 alpha*), et stimule le domaine de trimérisation du facteur de transcription HSF1, favorisant l'activation des gènes de réponse au choc thermique [61].

De plus, les études sur les miARN dans les espèces d'élevage sont de plus en plus nombreuses. Ces petits ARN non codants sont des régulateurs de la traduction capables de réprimer l'expression d'un gène : leur appariement à une séquence complémentaire de l'ARN messager du gène cible conduit à la répression traductionnelle ou à la dégradation de cet ARN messager. Plusieurs miARN sont identifiés chez le bovin, le mouton, le porc et le poulet comme ayant des rôles fonctionnels dans le tissu adipeux, les muscles squelettiques, les oocytes et durant le développement embryonnaire [62]. Des études sur le tissu adipeux chez les bovins ont montré que les profils d'expression des miRNA changent selon des facteurs environnementaux, comme par exemple l'épaisseur de graisse sous-cutanée et viscérale en fonction du régime alimentaire [63]. Les bovins ont

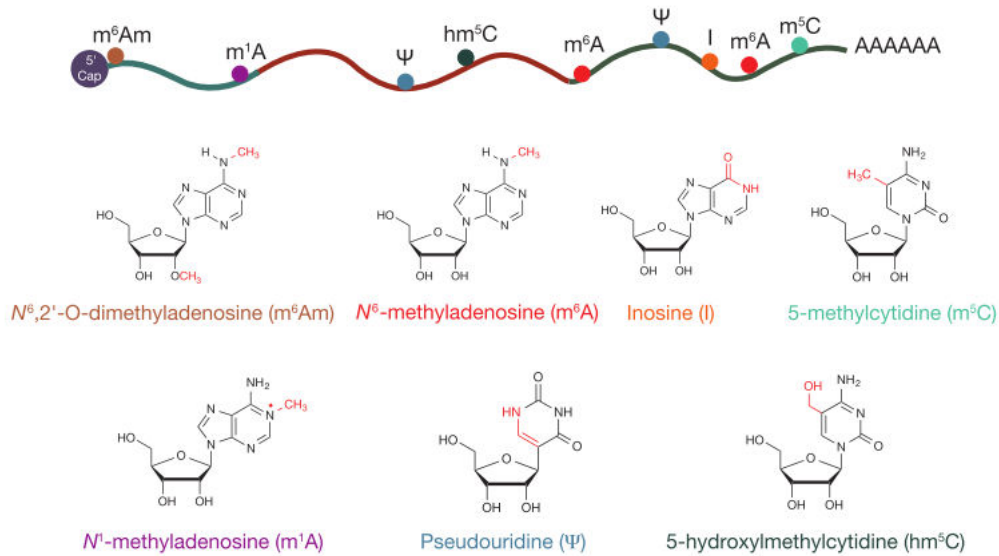


FIGURE 1.1 – Modification chimique des ARN messagers. Ces modifications constituent l'épitranscriptome, terme utilisé par analogie à la régulation épigénétique via la méthylation de l'ADN et les PTM des histones [64] (figure issue de [28]). Plus de 100 types de modifications post-transcriptionnelles ont été identifiées dans l'ARN. Ces modifications chimiques sont présentes dans les ARN non codants (ncRNA) et sont importantes pour le maintien des fonctions de traduction et d'épissage des ncRNA [65][66]. Plusieurs modifications dont la N^6 -méthyladenosine, la 5-méthylcytidine, l'inosine, la pseudouridine, la N^1 -méthyladenosine et la 5-hydroxyméthylcytidine sont retrouvées dans l'ARN messager des eucaryotes et peuvent influencer le métabolisme et la fonction de l'ARN messager.

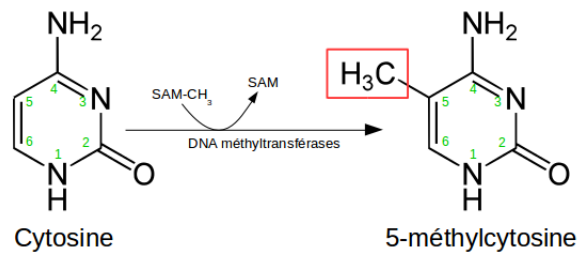


FIGURE 1.2 – Conversion d'une cytosine en 5-méthylcytosine. Cette conversion est catalysée par des DNA méthyltransférases. Transfert d'un groupement méthyle (CH_3) d'une S-adénosylméthionine (SAM) sur la position du 5^e carbone de la cytosine.

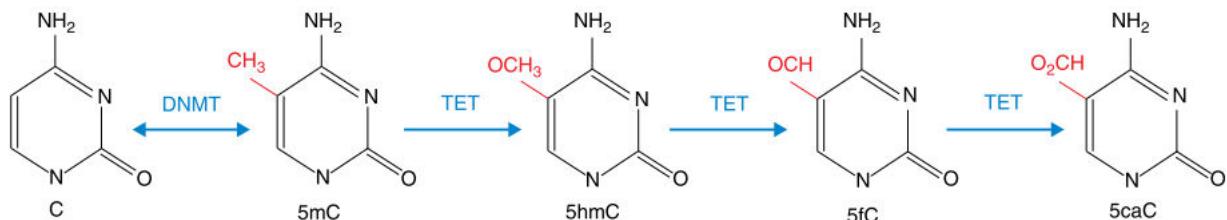


FIGURE 1.3 – Dérivés de l'oxydation de la 5-méthylcytosine. Ces réactions sont catalysées par les protéines de la famille TET (figure issue de [67]).

été nourris soit avec un régime riche en graisse (6% de graisse), soit avec un régime contrôle (2% de graisse). Un plus grand nombre de miRNA est détecté chez les animaux nourris avec le régime à haute teneur en graisse, ce qui suggère que ce régime augmente la complexité du rôle des miRNA dans la régulation de l'expression des gènes. De plus, les profils d'expression des miRNA diffèrent entre le dépôt de gras et le niveau de graisse selon le régime alimentaire, suggérant un rôle des miRNA dans la régulation de l'adipogenèse. Ainsi, les deux groupes d'animaux ont eu une réponse différente aux effets de l'environnement indiquant une interaction entre la régulation génétique et l'environnement.

Il existe également une régulation épigénétique des ARN qui modifie directement les ARN messagers (Figure 1.1).

1.1.4.3 Remaniement de la chromatine

Le remaniement de la chromatine implique le repositionnement ou la restructuration des nucléosomes dans la chromatine pour faciliter ou restreindre l'accessibilité à l'ADN. Ces modifications de conformation sont principalement effectuées par des complexes de remaniements de la chromatine ATP dépendants qui utilisent l'énergie de l'hydrolyse de l'ATP pour perturber les contacts entre l'ADN et les histones. Grâce à cette énergie, les nucléosomes glissent le long de l'ADN, ce qui soit altère leur espacement, soit les éjecte, ou encore remplace certaines de leurs histones [68]. La perturbation des interactions ADN-histones génère une boucle d'ADN. La propagation de cette boucle déplace l'ADN par rapport aux histones et permet le glissement des nucléosomes, favorisant ainsi la formation de la chromatine ouverte et l'activation transcriptionnelle (cf revue voir [69]). Pendant le développement embryonnaire, de nombreux remaniements de structure et d'assemblage de la chromatine apparaissent en réponse à des signaux extracellulaires. Ces remaniements dynamiques de la chromatine ont un rôle de régulation épigénétique dans plusieurs processus biologiques clés comme la réplication de l'ADN et la réparation, l'apoptose et la pluripotence [70]. Une régulation dysfonctionnelle de la chromatine a été observée dans de nombreuses maladies neuro-développementales et dans des incapacités mentales, associée à une régulation pathologique de gènes [69].

Cette dynamique de la chromatine est également liée à l'environnement (cf revue, voir [71]). Par exemple, la modification de la chromatine en réponse à un stress est impliquée dans le rôle de l'hormone ABA (*Abscisic Acid*) chez les plantes [72]. ABA favorise la dormance des graines et déclenche également un arrêt de croissance chez les embryons en post-germination qui sont soumis à un stress hydrique [73]. Cela s'accompagne d'une tolérance accrue à la dessiccation. La perte de fonction de la protéine de remaniement de la chromatine BRM (BRAHMA) provoque une hypersensibilité à l'hormone ABA pendant l'arrêt de la croissance en post-germination. De plus, la perte d'activité de BRM conduit à la déstabilisation d'un nucléosome, par phosphorylation, susceptible de réprimer la transcription de ABA. Ainsi, la régulation de ABA joue un rôle dans la croissance et la réponse à un stress hydrique chez les plantes, ce qui illustre un exemple de régulation du phénotype par l'environnement, via un remodelage de la chromatine.

1.2 Méthylation de l'ADN

Tous ces mécanismes épigénétiques démontrent que l'environnement module l'expression des gènes permettant ainsi une plasticité phénotypique [74]. Une des marques épigénétiques les plus étudiées est la méthylation de l'ADN. Chez les vertébrés, c'est une modification chimique qui peut affecter les bases cytosines dans les dinucléotides CG. La cytosine est une base pyrimidique de l'ADN. Elle peut porter un groupement méthyle CH_3 sur le carbone en position 5 entraînant l'apparition d'une 5-méthylcytosine (5mC, Figure 1.2). Cette modification de l'ADN s'effectue par les DNMT (*DNA methyltransferases*), famille d'enzymes impliquées dans l'établissement et la maintenance des motifs de méthylation de l'ADN (cf revue [75]). L'enzyme DNMT1 permet l'établissement des motifs de méthylation symétriques, en reproduisant la méthylation du brin d'origine sur le brin néosynthétisé, lors de la réplication de l'ADN [76]. De plus, le mécanisme de la méthylation de l'ADN peut être complété par des enzymes spécifiques comme la TET1 (*Ten-eleven translocation methylcytosine dioxygenase 1*) qui catalyse la conversion de la 5-méthylcytosine (5mC) en 5-hydroxyméthylcytosine (5hmC) [77][78]. La famille des protéines TET peut oxyder la 5mC pour générer ses dérivés : la 5hmC, la 5-formylcytosine (5fC) et la 5-carboxylcytosine (5caC) [79] (Figure 1.3). Le processus de conversion de la 5mC en 5hmC a été proposé comme la première étape de la déméthylation active de l'ADN chez les mammifères

[80]. Cependant, il est difficile d'analyser ces dérivés de la 5mC en raison de leur faible abondance dans le génome ($C \gg 5mC > 5hmC > 5fC > 5caC$ [81]).

La 5mC est prédominante chez les eucaryotes mais a pu être détectée chez les bactéries. Chez les procaryotes, la modification prédominante par la méthylation se fait par la N⁶-adénosine (6mA). De récentes études ont montré son rôle potentiel en tant que nouvelle marque épigénétique chez les eucaryotes, en plus de la 5mC [82]. Cependant, son abondance et son rôle chez les eucaryotes restent méconnus [83]. De même la méthylation de la cytosine en position 4 (4mC) a été retrouvée uniquement chez les génomes bactériens [84].

Trois contextes génomiques de la méthylation de la 5C existent chez les eucaryotes : CpG, CHG et CHH où H peut être A, T ou C. Chez les vertébrés, la majorité de la méthylation identifiée est en contexte CpG (*cytosine-phosphate-guanine*) à travers le génome. En revanche, les plantes et les invertébrés montrent un profil de méthylation plus en "mosaïque", c'est-à-dire que seuls des éléments génomiques spécifiques sont ciblés [85]. En effet, les séquences répétées et les séquences transcrits actives sont méthylées [86][87]. Dans le cas des séquences répétées, la méthylation induit une répression de la transcription et empêche l'expansion de ces éléments [88][89][90]. Chez les plantes, de petits ARN dérivés de transcrits répétés guident la méthylation *de novo* des éléments répétés [91]. De plus, la méthylation de l'ADN peut cibler les gènes transcrits et actifs. Ce processus a été montré dans beaucoup d'organismes avec une méthylation en mosaïque et peut même apparaître en l'absence de méthylation des éléments répétés chez *Ciona intestinalis* [86].

1.2.1 Îlots CpG et régulation de l'expression génique

Chez les vertébrés, la dispersion de la méthylation dans les génomes suggère que la méthylation est un état par défaut. La désamination spontanée d'une cytosine donne une uracile tandis que celle de la cytosine méthylée donne une thymine, base de l'ADN. Cette conversion peut être soit spontanée soit due à une désamination enzymatique. Dans le cas d'une désamination spontanée de C en U, la glycosylase uracile (UDG, *Uracil-DNA Glycosylase*) peut réparer la mutation grâce à un mécanisme d'excision / réparation de la base. En particulier, lors de la désamination d'une méthylcytosine (mC) en T, la thymine est réparée par les enzymes glycosylases TDG (*Thymine-DNA Glycosylase*) et MBD4 (*Methyl-CpG-Binding Domain protein 4*) qui réparent le mésappariement T/G sur le double brin d'ADN (appariement non "naturel" de l'ADN) pour éviter les mutations C/T [92]. Mais cette activité n'est pas totale. Des auteurs ont notamment montré que l'activité de la TDG était limitée, pour minimiser les excisions parasites des T dans les T/A, appariement "naturel" de l'ADN [93]. Même si elle est restreinte par ces mécanismes de réparation, la conversion de 5mC en T est donc existante dans le génome : la transition mC → T dans les dinucléotides CG méthylés est la mutation la plus fréquente observée dans les génomes, induisant parfois des maladies chez l'Homme [94], et chez les mammifères proches phylogénétiquement [95]. Cette diminution du nombre de CG au profit des TG est particulièrement importante dans les génomes de mammifères et des oiseaux mais moins évidente dans ceux des poissons ou des amphibiens [96]. Seules les régions génomiques appelées îlots CpG (CGI) sont exemptes de cette perte progressive des CG. Ce sont des regroupements de dinucléotides CG [97], principalement dans les régions promotrices du génome, et souvent peu méthylés. Selon Gardiner-Garden et Frommer [97], un îlot CpG est défini comme ayant un taux de C+G supérieur à 50%, un ratio de CpG observés sur CpG attendus de plus de 0,6 et un minimum de longueur de 200 pb. Leur distribution varie le long du génome selon les mammifères étudiés, et le taux de recombinaison et la taille des chromosomes peuvent avoir une influence sur l'évolution des CGI au cours de l'évolution des mammifères [98].

Une des premières études mettant en évidence des zones du génome riches en CpG et peu méthylées, chez la souris, a suggéré qu'ils pourraient être associés aux gènes [99]. Le faible taux de méthylation des dinucléotides CG dans la région promotrice des gènes est essentiel pour l'activité des gènes associés et a été mis en évidence, notamment, par la répression de la transcription causée par une méthylation artificielle de la région 5' du gène *aprt* (*adenine phosphoribosyltransferase*) de hamster, exprimé dans des cellules murines [100][101]. Des études ont également montré que l'intensité de la répression due à la méthylation dépend de la densité en CpG dans les régions promotrices, de sorte que la répression transcriptionnelle issue des CGI totalement méthylés ne peut même pas être réactivée par des *enhancers* (régions de l'ADN sur lesquelles se fixent les protéines pour amplifier la transcription) [102][103]. Il a été également montré que plusieurs facteurs de transcription se lient aux séquences méthylées pauvres en CG et mènent à la déméthylation [104][105][106]. Il est aussi probable que certains facteurs de transcription soient plus sensibles à la méthylation de l'ADN, en particulier ceux qui

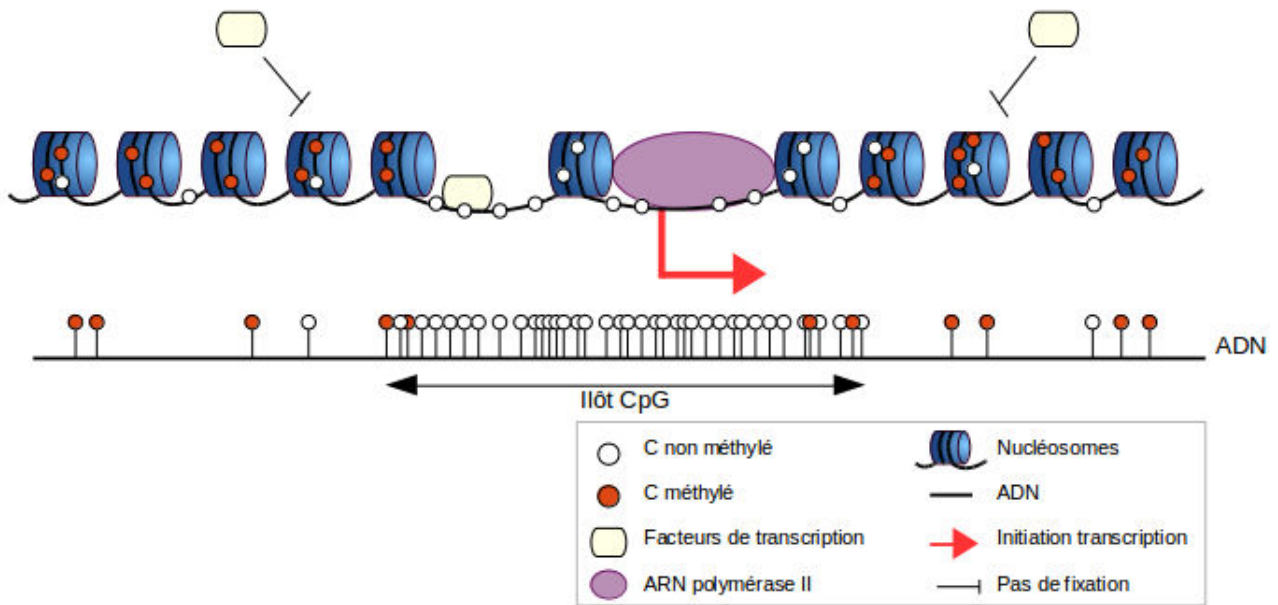


FIGURE 1.4 – Schéma des îlots CpG et de la régulation de l'expression associée. La séquence d'ADN contient des CG méthylés ou non. Le nombre des CG peut être particulièrement important dans certaines régions du génome et ainsi former des îlots CpG (CGI), majoritairement peu méthylés. Les CpG méthylés sont associés à de la chromatine stable et condensée, ce qui peut empêcher les facteurs de transcription de se fixer. L'ADN n'est alors pas accessible et la polymérase (ARN pol II) ne peut pas se fixer pour synthétiser l'ARNm. À l'inverse, une forte densité de CpG non méthylés est en général accompagnée par l'ouverture de la chromatine et permet aux facteurs de transcription de se fixer sur leur site de liaison, souvent riche en CG. Ils permettent de recruter l'ARN polymérase et ainsi initier la transcription.

contiennent un CG dans leur motif de liaison [107]. Les effets de cette sensibilité peuvent se retrouver dans les sites de liaison ayant une affinité plus faible dans lesquels la méthylation pourrait réduire davantage la probabilité de liaison. Cependant, ce rôle de la méthylation dans les régions pauvres en CG dans la régulation des gènes distaux reste méconnu.

À partir de la distribution des CG le long du génome, le sous-ensemble formé par les îlots CpG (CGI) constitue ainsi des régions génomiques d'intérêt. La grande majorité des CGI sont des sites d'initiation de la transcription, y compris en dehors des promoteurs annotés. Chez l'Homme, environ la moitié des CGI sont dans les promoteurs annotés [108]. Le reste des CGI est soit intra- soit intergénique et peuvent être appelés CGI "orphelins" [109]. Ce type de CGI peut aussi être associé à un site d'initiation de la transcription comme par exemple, le CGI dans l'intron 10 du gène *Kcnq1* (*Potassium Voltage-Gated Channel Subfamily Q Member 1*) soumis à empreinte qui a été identifié comme à l'origine d'un ARN long non codant nécessaire pour l'empreinte génomique parentale de plusieurs gènes [110]. Ces CGI orphelins, détectés dans les régions hors promoteurs annotés, peuvent également représenter de nouveaux promoteurs, pas encore identifiés [109]. Ils sont souvent marqués par la triméthylation de l'histone 3 (H3K4me3), qui est une signature des promoteurs actifs. Beaucoup de gènes proches de ces CGI orphelins sont actifs de façon tissu spécifique [109], ce qui suggère qu'ils sont fortement régulés. Ces CGI sont probablement des promoteurs alternatifs aux gènes annotés proches [111].

Il est maintenant connu que la méthylation des CGI peut entraîner une répression stable d'un gène [108] (Figure 1.4). Peu de CGI changent de statut de méthylation pendant un développement normal. Par exemple, un ensemble de gènes spécifiques de lignée germinale requièrent une méthylation du promoteur pour leur répression dans les cellules somatiques [112]. Cependant, des études ont montré que la méthylation du gène peut mener à l'activation de la transcription [113]. En effet, il a été montré que des régions transcrites des gènes sont très méthylées et que le niveau de méthylation est en général corrélé de façon positive avec le niveau d'expression [114][115]. La méthylation de l'ADN dans les régions transcrites pourrait potentiellement interagir avec d'autres éléments fonctionnels pour maintenir l'efficacité de la transcription [116][111][117][118]. De plus, la méthylation de l'ADN n'est pas seulement associée à la régulation de l'expression des gènes (répression et activation) : elle est un phénomène complexe et dynamique qui peut réguler l'épissage, le positionnement des nucléosomes et le recrutement des facteurs de transcription [119].

1.2.2 Méthylation de l'ADN pendant le développement

Les niveaux de méthylation de certains gènes varient entre différents types cellulaires [120]. Des marques spécifiques sont mises en place lors du développement et permettent une expression tissu-spécifique des gènes concernés. Ainsi, même si le génome est identique dans toutes les cellules, le taux de méthylation varie dans ces gènes selon les différents types cellulaires [30].

Il est maintenant connu que des changements anormaux d'états de méthylation pendant la période de développement embryonnaire peuvent entraîner des conséquences sur le phénotype, dont l'apparition de maladies chez l'Homme [121][122]. Pourtant, malgré l'importance démontrée de son rôle dans la régulation de l'expression des gènes, en particulier au cours de l'embryogenèse, les mécanismes moléculaires sous-jacents restent peu connus chez le poulet. Durant le développement, la dynamique de la méthylation de l'ADN dans l'embryon de poule comprend plusieurs périodes critiques et est sensible aux stimuli environnementaux. En effet, Li et al ont montré que la méthylation de l'ADN varie pendant le développement embryonnaire des poulets de chair [120]. Globalement, la méthylation de l'ADN dans les embryons augmente au fur et à mesure du développement dans le cœur, le foie et le muscle. Dans le muscle et le foie, la méthylation du promoteur du gène *TNF- α* (*Tumor Necrosis Factor alpha*), gène impliqué dans l'immunité, augmente jusqu'au 14^e jour de développement, puis diminue à partir du 17^e jour. La méthylation du corps du gène reste faible (inférieure à 4%) au début du développement avant d'augmenter dès le 17^e jour. Le niveau d'expression de *TNF- α* augmente principalement à partir du 17^e jour de développement. Dans le gène *IGF2* (*Insulin-like Growth Factor 2*), gène ayant un rôle dans la prolifération et la division cellulaire, la méthylation du promoteur diminue à partir du 17^e jour de développement, et la méthylation du corps du gène augmente continuellement lors du développement. Son expression augmente pendant les phases embryonnaires. De plus, les niveaux d'ARNm augmentent avec la succession des étapes de formation de l'embryon. Ces résultats montrent que les niveaux de méthylation des promoteurs des gènes étudiés dans les différents tissus sont liés aux niveaux d'expression des gènes. L'expression génique peut être régulée simultanément par l'hypométhylation du promoteur et l'hyperméthylation du

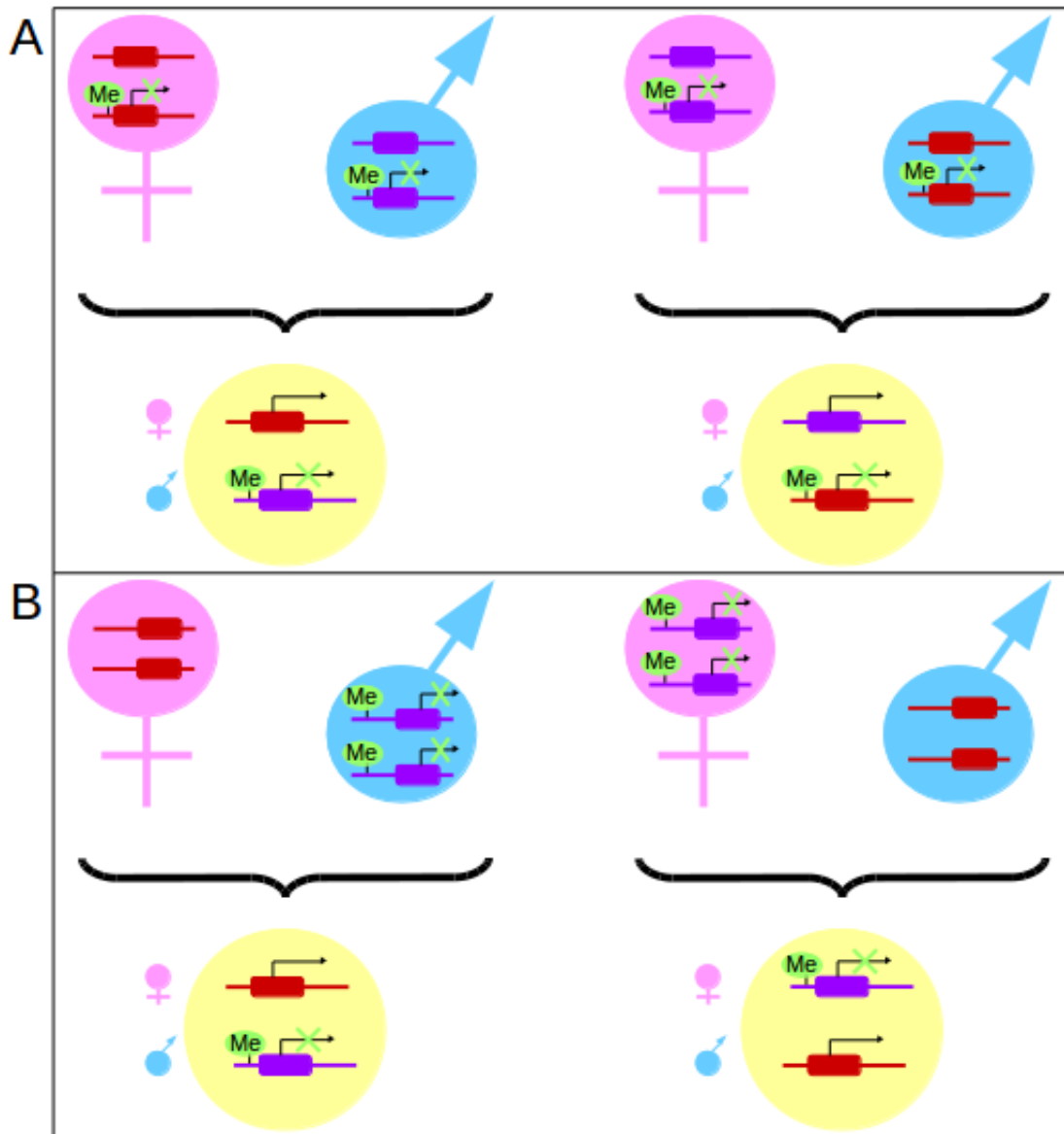


FIGURE 1.5 – Schéma du phénomène de méthylation allèle-spécifique. La méthylation allèle-spécifique (ASM) peut être observée soit dans les mécanismes dépendants de l'origine parentale, comme le phénomène d'empreinte génomique parentale (A), soit dans les mécanismes dépendants de la séquence (B). Le premier croisement (A, à gauche) illustrant le phénomène d'empreinte génomique parentale montre que l'allèle porté par la mère (en rouge) est non méthylé chez le descendant (en jaune) et l'allèle du père (en violet) est méthylé. Dans le second croisement (A, à droite) illustrant l'empreinte génomique parentale, c'est l'allèle maternel (violet) qui est non méthylé et l'allèle paternel (rouge) qui est méthylé : la méthylation dépend de l'origine parentale et non de la séquence de l'ADN. Dans cet exemple, lors de la transmission des allèles au descendant, l'allèle exprimé est toujours celui transmis par la mère, quel qu'il soit. Au sein des deux croisements qui représentent la méthylation allèle spécifique de la séquence (B), la méthylation de l'allèle violet est dépendante de la séquence de l'allèle, indépendamment de l'origine parentale.

corps du gène. Ainsi, lors du développement embryonnaire, l'expression des gènes spécifiques de tissus ou de cellules peuvent être corrélés à l'altération de marques épigénétiques.

1.2.3 Héritabilité et transmission

Les marques épigénétiques peuvent être transmises par la mitose à travers des générations cellulaires et par la méiose à travers une succession de plusieurs générations [123]. Des études ont montré que la transmission potentielle des modifications épigénétiques aux générations suivantes serait susceptibles d'influencer les phénotypes adultes associés [123][32]. Chez les mammifères, la transmission des effets des variations environnementales peut se faire soit directement durant le développement embryonnaire pendant la gestation via l'environnement maternel, soit durant la vie adulte où les gamètes produites peuvent transmettre des modifications épigénétiques à la descendance [124]. En effet, lorsqu'une femelle gestante F0 est exposée à des perturbateurs épigénétiques, les lignées somatiques et germinales du fœtus F1 sont également directement exposées. Ces cellules germinales exposées de F1 sont à l'origine de l'individu F2. C'est seulement à partir de la génération suivante, F3, que des individus non directement exposés aux variations subies par la femelle F0 pendant la gestation sont observés. Les effets observés en F1 sont qualifiés de multigénérationnels si ils atteignent F1 et F2. Cependant, si des effets en F3 sont remarqués alors on parle d'effets transgénérationnels car il y a une transmission épigénétique au cours des générations (F3 n'étant pas exposé de façon directe aux modifications épigénétiques). Un exemple connu est la famine hollandaise qui a sévi en 1944. Les mères exposées à la famine pendant la grossesse ont donné naissance à des enfants plus sensibles à des problèmes de santé (maladies cardiaques et métaboliques) en comparaison aux enfants nés l'année suivante [125]. Parmi les effets épigénétiques observés, IGF2, impliqué dans la croissance, est hypométhylé chez les enfants dont les mères ont été exposées à la famine [122]. Cela souligne que dans le développement embryonnaire des mammifères, la maintenance et l'établissement des marques épigénétiques sont essentiels [126]. Durant le développement embryonnaire, la méthylation des génomes parentaux présente dans les cellules germinales à l'origine des embryons est effacée et est remplacée par des marques embryonnaires importantes pour le développement et pour la pluripotence cellulaire. La méthylation établit à nouveau des motifs de méthylation dans les cellules somatiques pendant la phase d'implantation (phénomène de reprogrammation [126]).

Chez les oiseaux, quelques études suspectent des phénomènes épigénétiques transgénérationnels mais aucune n'a validé cette hypothèse. Chez la caille, il a été suggéré que l'environnement embryonnaire affecte le phénotype de la descendance sur trois générations [127]. Sur deux lignées de caille croisées en miroir, une lignée a subi une injection de génistéine dans l'œuf, phyto-œstrogène impliqué dans la modification de la méthylation de l'ADN. À la troisième génération, les femelles issues de la lignée injectée pondent des œufs plus tardivement que les femelles de la lignée contrôle. Le poids du corps des adultes et le comportement social sont également impactés par le traitement à la génistéine. Néanmoins, aucune différence de méthylation globale entre les deux lignées n'a été observée sur des échantillons sanguins. Une analyse multigénérationnelle, chez le poulet, a montré qu'une altération du régime de la mère peut mener à des variations épigénétiques chez les descendants [128]. Le régime des mères a été enrichi par du zinc ce qui a des conséquences physiologiques sur l'intestin des descendants (augmentation de la hauteur des villosités et diminution de la profondeur des cryptes). Chez les poules supplémentées en zinc, l'hypométhylation de l'ADN et l'hyperacétylation de la lysine 9 de l'histone H3 ont été observées dans la région promotrice du gène A20 (ou TNFAIP3, *Tumor Necrosis Factor Alpha Induced Protein 3*) entraînant une activation de l'expression du gène A20 impliqué dans la réponse anti-inflammatoire. Une modification de la composition de l'alimentation peut entraîner une diminution de la réponse inflammatoire, via des modifications épigénétiques du promoteur A20.

1.2.4 Méthylation allèle-spécifique

Comme décrit ci-dessus, la méthylation de l'ADN est une marque épigénétique qui est impliquée dans la régulation de la transcription. Les motifs de méthylation sont spécifiques aux tissus. De plus, la méthylation de l'ADN peut aussi être spécifique selon les allèles au sein d'une cellule et être associée à une expression spécifique d'allèle. La méthylation allèle-spécifique (ASM, *Allele-specific Methylation*) est un événement où la méthylation d'un allèle diffère de l'autre allèle. Ce phénomène a été observé soit dans les mécanismes dépendant de l'origine parentale comme les gènes soumis à l'empreinte parentale (Figure 1.5 A) et l'inactivation du chromosome X chez les femelles, soit dans les mécanismes dépendant de la séquence (Figure 1.5 B) [129][130].

Une proportion significative de la variance phénotypique peut être expliquée par des loci soumis à empreinte. À ce jour, il a été montré que l'empreinte génomique parentale n'existe probablement pas chez les oiseaux [131][132][133]. Le mécanisme d'inactivation du chromosome X est développé ci-dessous (cf partie 1.2.5 page 39).

En dehors de l'empreinte parentale, l'ASM est dépendante de la présence de SNP hétérozygotes qui perturbent le potentiel de méthylation des CG [134]. En effet, les SNP peuvent affecter les niveaux de méthylation des sites CpG voisins. Chez la souris, l'ASM dépendante de la séquence aux loci non soumis à empreinte est plus fréquente le long du génome que le phénomène d'empreinte [135]. Cette étude a montré que l'ASM est principalement trouvée dans les régions intergénomiques et introniques, et est réduite au niveau des promoteurs proximaux. Cela est dû, entre autres, à la densité des SNP. En effet, la fréquence des SNP associés à une méthylation spécifique de l'allèle est élevée. Néanmoins, les gènes où peu d'ASM est détectée sont principalement les gènes codants pour les gènes à homéoboîte (ou *homeobox*) (impliqués dans le développement embryonnaire), les facteurs de transcription et les régulateurs du développement [135]. Cela suggère que la méthylation des promoteurs des gènes associés au développement et des gènes de ménage est soumise à une régulation stricte, peu susceptible d'être perturbée.

La majorité de l'ASM est associée avec des événements génétiques régulés en *cis* (mécanisme intervenant sur le même chromosome). Par exemple, chez *Nasonia*, insecte parasite, les gènes détectés différenciellement méthylés entre deux espèces sont aussi différenciellement méthylés entre les deux allèles dans la descendance issue du croisement entre ces deux espèces, indiquant ainsi la présence d'éléments en *cis* qui ciblent la méthylation du gène [136][137]. Comme chez la poule, aucune empreinte génomique parentale ou région différenciellement méthylée d'origine parentale n'a été trouvée chez *Nasonia*. Cela, combiné avec l'héritabilité très stable de la méthylation observée entre les générations, suggère soit un motif de régulation en *cis* pour la méthylation au niveau de l'ADN soit une héritabilité d'un signal épigénétique.

De plus, l'ASM apparaît avoir un rôle dans la régulation de l'expression allèle-spécifique (ASE) des gènes non soumis à empreinte. Une grande proportion de phénomènes ASE sont enrichis aux alentours des phénomènes ASM où un allèle hypométhylé correspond à un allèle très exprimé [138]. Chez *Nasonia*, une corrélation positive entre l'ASM et l'ASE a été détectée [136]. L'impact de l'ASM dans les changements d'expression des gènes à proximité a donc le potentiel d'affecter la variation phénotypique et dans certains cas, d'induire des maladies [139]. L'ASM n'a jamais été mise en évidence chez la poule.

1.2.5 Compensation de dose

Le déterminisme sexuel de nombreuses espèces est génétiquement lié à la paire des chromosomes sexuels [140]. Deux systèmes existent : le système ZW (♀ hétérogamétique ZW et ♂ homogamétique ZZ) est retrouvé principalement chez les oiseaux, de nombreux serpents et les papillons, tandis que le système XY (♀ homogamétique XX et ♂ hétérogamétique XY) est présent chez les mammifères, coléoptères et les drosophiles [141]. La monosomie du chromosome X pour le système XY et du chromosome Z pour le système ZW provoque, pour les gènes liés à l'X ou au Z, une diminution de la dose des gènes dans le sexe hétérogamétique. Un mécanisme de compensation de dose peut permettre d'équilibrer l'expression des gènes entre les sexes. Chez les mammifères, cette compensation de dose est obtenue par l'inactivation de l'un des deux chromosomes X dans le sexe homogamétique (femelles XX), ce qui permet des niveaux d'expression similaires entre mâles et femelles des gènes liés à l'X [142][143]. Cependant, il y a une variation importante dans ce phénomène et dans l'intégralité de la compensation de dose entre espèces. Chez les oiseaux, cette compensation de dose est incomplète [144] : la plupart des gènes liés au Z montrent des ratios d'expression qui varient entre mâles et femelles, la grande majorité étant plus exprimés chez le mâle, ce qui souligne qu'un mécanisme de compensation de dose apparaît mais est toujours incomplet [141][145]. La variabilité dans ces ratios d'expression indique un ensemble complexe de mécanismes qui détermine la compensation de dose [146][141]. Même si les mécanismes épigénétiques sous-jacents liés à la compensation de dose restent peu étudiés chez les oiseaux, un exemple connu est la région MHM (*Male Hypermethylated*) [147][148]. C'est un locus du chromosome Z qui est fortement méthylé et dont l'expression est réprimée chez les mâles. Chez les femelles, ce locus MHM est hypométhylé et transcrit en ARN long non codant. La chromatine à ce locus est fortement enrichie par une hyperacétylation de l'histone 4 sur la lysine 16 (H4K16ac), mécanisme spécifique chez la femelle [149].

1.2.6 Phénomènes épigénétiques chez les oiseaux : la revue

Une revue a été réalisée pour synthétiser l'apport des technologies haut-débit dans l'état actuel des connaissances des mécanismes épigénétiques décrits précédemment, en particulier la méthylation de l'ADN et les modifications d'histones, chez la poule. L'avancée majeure des technologies de séquençage, ces dernières années, a permis de caractériser plus finement le paysage de la méthylation de l'ADN et des marques d'histones le long des génomes, et d'identifier des profils caractéristiques pour les épigénomes. L'apport croissant de données issues des nouvelles technologies à haut-débit favorise une meilleure compréhension des mécanismes moléculaires sous-jacents. De plus, la revue a été mise à jour et complétée pour être publiée, en français, dans la revue *Productions Animales* de l'INRA (Annexe A, le manuscrit est accepté avec révisions mineures).

Malgré une redondance partielle entre l'introduction de thèse et la revue, celle-ci complète l'introduction de thèse sur les phénomènes épigénétiques (cf partie 1.1.4 page 27), généraliste et prenant des exemples chez les mammifères, par une synthèse des connaissances épigénétiques chez la poule. Les techniques de biologie moléculaire et les outils d'analyses utilisés dans la thèse et évoqués dans la revue seront décrit plus en détails dans la partie suivante (cf partie 1.3 page 61).

Review

Genome-Wide Epigenetic Studies in Chicken: A Review

Sarah-Anne David ^{1,†}, Marjorie Mersch ^{2,†}, Sylvain Foissac ², Anne Collin ¹,
Frédérique Pitel ^{2,*}  and Vincent Coustham ^{1,*} 

¹ URA, INRA, 37380 Nouzilly, France; sarah-anne.david@inra.fr (S.-A.D.); anne.collin@inra.fr (A.C.)

² GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet Tolosan, France; marjorie.mersch@inra.fr (M.M.); sylvain.foissac@inra.fr (S.F.)

* Correspondence: frederique.pitel@inra.fr (F.P.); vincent.coustham@inra.fr (V.C.); Tel.: +33-247-427-923 (F.P. & V.C.)

† These authors contributed equally to this work.

Academic Editor: Muller Fabbri

Received: 26 September 2017; Accepted: 21 November 2017; Published: 23 November 2017

Abstract: Over the years, farmed birds have been selected on various performance traits mainly through genetic selection. However, many studies have shown that genetics may not be the sole contributor to phenotypic plasticity. Gene expression programs can be influenced by environmentally induced epigenetic changes that may alter the phenotypes of the developing animals. Recently, high-throughput sequencing techniques became sufficiently affordable thanks to technological advances to study whole epigenetic landscapes in model plants and animals. In birds, a growing number of studies recently took advantage of these techniques to gain insights into the epigenetic mechanisms of gene regulation in processes such as immunity or environmental adaptation. Here, we review the current gain of knowledge on the chicken epigenome made possible by recent advances in high-throughput sequencing techniques by focusing on the two most studied epigenetic modifications, DNA methylation and histone post-translational modifications. We discuss and provide insights about designing and performing analyses to further explore avian epigenomes. A better understanding of the molecular mechanisms underlying the epigenetic regulation of gene expression in relation to bird phenotypes may provide new knowledge and markers that should undoubtedly contribute to a sustainable poultry production.

Keywords: high-throughput sequencing; epigenomics; poultry; bird; chicken; DNA methylation; histone post-translational modifications

1. Introduction

Beyond its well-established role as a developmental model for embryogenesis studies, poultry is a major source of healthy proteins through meat and egg production. Over the past 50 years, studies have focused on the improvement of poultry production notably through genetic selection. However, quantitative trait analyses and genome-wide association studies have shown that genetic variation is not the only contributor to phenotypic variability [1]. An increasing number of evidence suggests that epigenetics may also contribute to the phenotypic diversity of animals, in interaction with their environment [2]. Recently, epigenetic reprogramming of gene expression was shown to impact, in the long term, the phenotypes of birds. Notably, several studies have investigated epigenetic marks that may be implicated in behavior [3,4], adaptation [5,6] and the impact of nutrition on chicken performance [7,8]. Moreover, the available information on chicken epigenetics is growing thanks to the recent advances in high-throughput sequencing techniques (HTS). To date, only one epigenome-wide analysis was performed in a non-chicken organism, the Great Tit [9]. Here, we review

the recent gain of knowledge on the chicken epigenome by focusing on two epigenetic marks that have been extensively studied during the past decade, DNA methylation and histone post-translational modifications (HPTM), keeping in mind that other epigenetic mechanisms such as non-coding RNAs or mRNA post-translational modifications may also play a major role in phenotypic variability [10,11]. Moreover, DNA methylation and HPTM studies confirmed that these marks were similarly distributed in the chicken genome compared to mammals, and that they displayed similar patterns of association with gene expression status. We then describe examples of global changes in DNA methylation and histone marks that contribute to biological processes such as resistance to infection, or the dynamics of these marks during the lifespan of laying hens. We finally discuss the quality of genome annotation, bioinformatics tools and computational skills that are required for such genome-wide analyses.

2. From Epigenetics to Epigenomics

2.1. Epigenetic Marks

Waddington defined an epigenetic trait in the 1950s as “a trait with a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” [12]. Nowadays, epigenetics is commonly defined as the molecular mechanisms involved in the regulation of gene expression that are reversible and heritable (by mitosis and potentially meiosis) without alteration of the DNA sequence. It was shown that epigenetic modifications in eukaryotes can impact gene expression under the influence of the environment (for a review see [13]).

Originally, the most studied epigenetic mark associated with major changes in gene expression was DNA methylation [14,15]. In animals, DNA methylation mostly targets cytosine residues at the carbon 5 position (5mC, [16]), although methylation of adenosine residues has also been reported in mammalian embryonic stem cells [17]. DNA methylation occurs in three different sequence contexts: CpG, CHG or CHH (where H is A, T or C). In vertebrates, 5mC predominates in the CpG dinucleotides. The distribution of 5mC across the genome is non-random and CpG dinucleotides tend to be frequently concentrated in CpG islands [16]. The methylation of DNA in gene promoters is usually negatively associated with gene expression [18–21], whereas it may be positively correlated with transcription within the bodies of actively transcribed genes [22]. In human monozygous twins, differential methylation was shown across several tissues, suggesting that the life history of an individual is able to modify the way its genome is expressed through DNA methylation [23]. The characterization of DNA methylation associated with phenotypic variation is documented in several species [13]. Notably, a diet supplemented with methyl donors (such as folic acid, vitamin B12, betaine and choline) resulted in altered offspring phenotypes in the gestating viable yellow agouti (A^{vy}) mouse model, explained by the differential methylation of a retrotransposon [24].

Another type of epigenetic mark of interest is histone post-translational modification. Histones are proteins found in eukaryotes that package DNA into structural units called nucleosomes. The histones H2A, H2B, H3 and H4 form a core wrapped by 147 base pairs of DNA [25]. Amino acid residues of histones (tails or core) can be targeted by covalent modifications influencing chromatin structure [26,27]. These HPTM, such as methylation, acetylation, ubiquitination, phosphorylation and sumoylation [28], can influence chromatin conformation, recruitment or eviction of interacting proteins and transcription [29]. They can either be associated with active or repressed gene expression. Acetylation of lysine 27 on histone H3 (H3K27ac) and tri-methylation of lysine 4 on histone H3 (H3K4me3) are commonly categorized as “active” marks. On the one hand, acetylation of positively charged lysine residues on histone tails was shown to neutralize their charge, resulting in a lower interaction between the histone tail and the nucleosome, therefore opening the chromatin (for review see [30]). On the other hand, H3K4me3 is frequently associated with the start of genes and favors an “open” chromatin state enabling gene expression (for review see [31]). In *Arabidopsis* plants, following heat stress, a long lasting accumulation of H3K4me3 was shown at the locus of highly expressed genes implicated in heat stress memory [32]. In contrast, trimethylation is commonly associated with gene repression when

it targets lysine 9, 36 and 27 of histone H3 (H3K9me3, H3K36me3, H3K27me3). Unlike H3K4me3, H3K27me3 is often associated with a “closed” chromatin state linked to the memory of gene repression (for review see [31]). A well-described illustration of the role of H3K27me3 in epigenetic control of phenotype through gene regulation is vernalization in *Arabidopsis* [33]. Vernalization promotes flowering through the cold-induced repression of a floral repressor which is maintained by the spreading of H3K27me3 marks all over the *flowering locus C (FLC)* locus. So far, H3K4me3 and H3K27me3 are the two most studied HPTM. However, the genome-wide distributions of other marks such as H3K9me3, H3K36me3 and H3K27ac have also been investigated in several species [34]. At a finer resolution, combinatorial models of various HPTMs were shown to define specific chromatin states (for a review see [35]) and predict gene expression levels [36]. Thus, with the development of HTS and careful experimental design, the characterization of epigenomes should allow the investigation of interactions between particular chromatin states and the associated phenotypic impact.

2.2. Epigenomics Methods

For the past two decades, the advent of HTS allowed researchers to extend the knowledge of different epigenetic marks at the whole genome level and their interplay in several tissues. Several consortia emerged such as the ENCODE project (Encyclopedia of DNA Elements, [37]), the National Institutes of Health (NIH) Roadmap Epigenomics Project Mapping Consortium [38], the International Human Epigenome Consortium [39] and modENCODE (Model Organism Encyclopedia of DNA Elements [40]). The goal of these consortia was to streamline experimental design and analyses, to provide guidelines and to coordinate the sequencing efforts. They produced several databases containing a collection of epigenomes (mostly based on mammalian species). Since 2015, the willingness to better understand genetic, epigenetic and phenotypic diversity of domesticated species such as the chicken has driven to create a consortium called FAANG (Functional Annotation of Animal Genomes, [41]). Beyond its goal of providing high quality functional annotation of animal genomes including a comprehensive description of the epigenome landscape, FAANG also aims to contribute to the medical and agricultural research (production and welfare) by standardizing core assays and experimental protocols, coordinating and facilitating data sharing as well as establishing an infrastructure to analyze these data.

DNA methylation can be investigated by several approaches [42]. The gold standard method to date is the bisulfite treatment method followed by bisulfite sequencing (BS-seq; Figure 1). This approach is used to analyze methylation patterns at a single base pair resolution. Genomic DNA is treated with sodium bisulfite that deaminates cytosine to uracil. Deamination is blocked by the presence of methylation that can then be inferred by the identification of unconverted C residues [43]. This base modification can be identified by HTS technologies by comparing the sequences of generated reads (for a review see [44]) with a genomic reference allowing to identify the original methylation status. The entire genome of different organisms was initially sequenced using BS-seq (whole genome bisulfite sequencing (WGBS) or MethylC-seq, [45]). More recently, reduced representation bisulfite sequencing (RRBS), presented as a more cost-efficient method, was developed to only sequence the fraction of the genome that is enriched in CpG sequences, i.e., mostly CpG islands [46,47]. This technique combines digestion by restriction enzyme(s) and sequencing in order to provide high sensitivity for areas with high CpG content [48]. Although these two techniques are based on the same principle, WGBS is recommended for the detection of widespread CpG methylations on the genome, while RRBS is more specific to CpG islands, and therefore, promoter regions [49]. Other techniques can be used to analyze DNA methylation: methyl-sensitive restriction enzyme digestion followed by sequencing (MRE-seq), affinity enrichment of methylated DNA fragments (methylated DNA immunoprecipitation sequencing, MeDIP-seq) and methyl-binding domain sequencing, MBD-seq [50]. In particular, MeDIP-seq was a popular alternative to bisulfite sequencing before sequencing costs significantly dropped, as it combines immunoprecipitation of methylated genomic DNA with an antibody specific to 5mC followed by high throughput sequencing [51]. However, these methods present limitations compared to the single

base pair resolution obtained by bisulfite techniques. For MRE-seq, the resolution depends on the frequency of the enzyme cut sites. For MeDIP-seq, the sequencing resolution is linked to the size of DNA fragments, CpG density as well as the specificity and affinity of the antibody. Another new technique combining genotyping by sequencing [52] with MeDIP was created to reveal methylation patterns on a reduced fraction of the genome [53]. This approach is based on genome digestion with enzymes unrelated to CpG islands that prevents bias toward CpG islands.

Recently, 5-hydroxymethylcytosine (5hmC) has gained attention because of its function in DNA demethylation and regulation of tissue-specific gene expression [54–56]. The bisulfite treatment does not allow to distinguish 5mC from 5hmC, but the impact of 5hmC on 5mC data interpretation should be minimal due to its low frequency in many mammalian tissues [57]. A strategy to study 5hmC is to obtain specific 5mC signal by performing an oxidation step prior to bisulfite sequencing and by subtracting this signal to the actual bisulfite signal obtained from a non-modified template (corresponding to 5mC + 5hmC) [58,59]. It is worth noting that the oxidation step is not routinely performed to accurately study 5mC profiles as oxidation may damage DNA during the process and because 5hmC is of very low abundance in most tissues.

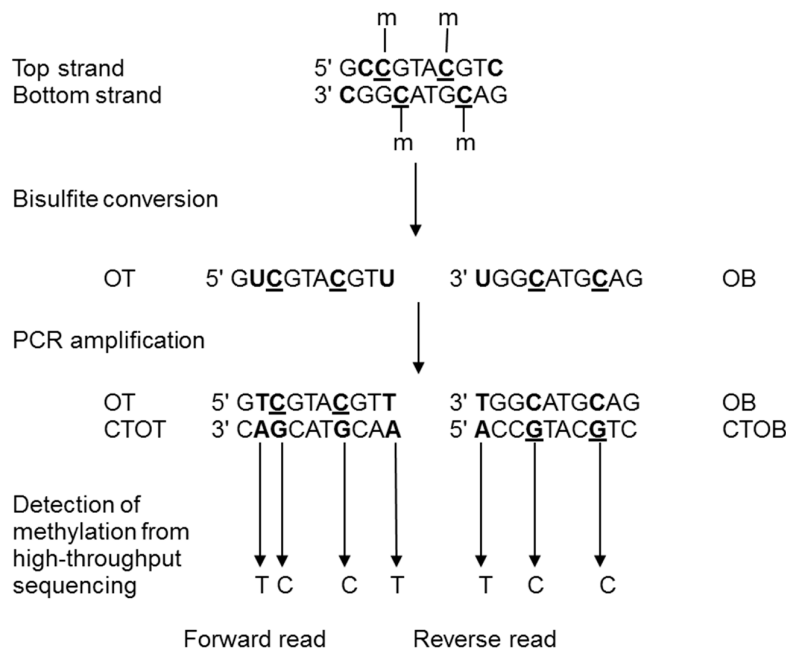


Figure 1. Principle of DNA methylation analysis by bisulfite treatment (modified from [60]). On a DNA sequence, unmethylated cytosines are converted to uracil by the bisulfite treatment and to thymines after PCR amplification. After sequencing, the level of methylation is detected by counting cytosines and thymines for each position. m, methyl group on cytosine; OT, Original Top strand; CTOT, strand complementary to the original top strand; OB, original bottom strand; and CTOB, strand complementary to the original bottom strand.

To study histone marks, the gold-standard technique is chromatin immunoprecipitation followed by HTS (ChIP-seq). This technique maps genome-wide DNA fragments associated with proteins using specific antibodies [61–63]. Therefore, by using an antibody that is specific to a particular HPTM residue, ChIP-seq allows to map the DNA sequences that are bound to the nucleosomes bearing the mark of interest. Many factors can be profiled with ChIP-seq, such as transcription factors, chromatin binding proteins and histone modifications, provided that a highly-specific antibody is available (for an assessment of HPTM-directed antibodies for ChIP, see [64]). To perform ChIP-seq, chromatin can either be cross-linked to covalently fix the interactions between the DNA and the protein or mark of

interest (X-ChIP-seq), or used in the native state after digestion using frequent cutter enzymes such as MNase (N-ChIP-seq, [65]). Immunoprecipitated DNA fragments are then subjected to sequencing (Figure 2). Recently, ChIP-seq was improved by including an enzymatic digestion to retrieve smaller DNA fragments associated with the MPTH in order to increase the resolution and to limit non-specific DNA sequencing (ChIP-exo, [66]).

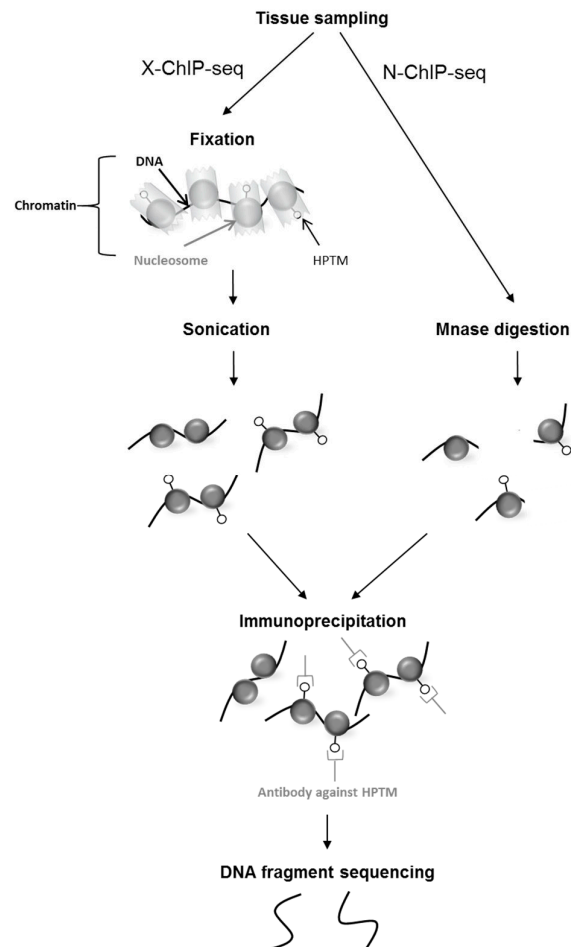


Figure 2. Protocol of chromatin immunoprecipitation followed by high throughput sequencing. The chromatin is extracted from tissues and can be fixed (X-ChIP-seq) or left unfixed (N-ChIP-seq). For X-ChIP-seq the interaction between the DNA and histone bearing the mark is fixed and the chromatin is shredded by sonication to obtain fragments between 100 bp and 600 bp. For N-ChIP-seq, native chromatin is digested by an endonuclease (MNase) that specifically cuts unprotected DNA between nucleosomes to obtain mono-nucleosomes. The chromatin is fragmented by mechanical or enzymatic treatments and immunoprecipitated with an antibody directed specifically against the histone post-translational modifications (HPTM) of interest. After several washes, the DNA is eluted and sequenced following library creation.

3. New Knowledge Gained on Chicken Models by High-Throughput Sequencing

In this review, we focus on the chicken genome that was the first genome sequenced for bird species. Chicken is a commonly used model to study developmental biology (in particular during embryogenesis), immunology and microbiology. Chicken is also an appropriate model to improve the mammalian genome annotation [67]. Indeed, long blocks of conserved synteny between chicken

and human genomes could be explained by common intra-chromosomal rearrangements and few chromosomal translocations.

3.1. The Chicken Genome

The chicken genome is about 2.5 times smaller than the human genome (1200 Mb vs. 3200 Mb). It is composed of 38 pairs of autosomes and two sex chromosomes (Z and W). Many of the autosomes (29 pairs in total) are relatively small and uniform in size, and are termed 'microchromosomes'. About 38% of CpG islands are conserved between human and chicken [67]. In the chicken genome, CpG islands are mostly localized on microchromosomes which have a high density of genes [68]. The *Gallus gallus* genome assembly was obtained from a single female red jungle fowl chick. The first two versions were assembled using a whole-genome Sanger sequencing [67,69]. The latest version of the chicken genome assembly (*Gallus_gallus*-5.0; GCA_000002315.3), of 1.2 Gb size, was released in 2015 and was obtained through the combination of short read (Illumina) and long read technologies (PacBio), which significantly improved the contig length and thus the quality of the assembly [70,71]. In the latest version of the chicken genome assembly, the number of assembled chromosomes was improved (from the autosomes GGA1-28 and GGA32 to autosomes GGA1-28 and GGA30-33, [70]). The increased coverage (from 12X to 70X) allowed a more complete chicken genome assembly, but the large amount of data produced also increased the number of scaffolds or unplaced contigs. About 183 Mb of new sequences were added, networks were improved, errors identified and a more complete set of genes was determined [70].

HTS technology has also been used to successfully sequence species such as Turkey which was the second avian genome sequenced in 2010 [67,72]. At the moment, the genomes of nearly 60 different bird species have been sequenced [73] and more are planned to complete the set of available avian genomes. The current limitation in the study of these genomes is the lack of assembly. Indeed, as of today, only seven genomes have been assembled and annotated: the chicken (fifth version), the turkey (second version), the great tit (second version), the zebra finch (second version), the quail (second version), the helmeted guinea fowl (first version) and the collared flycatcher (first version; source: NCBI). Other genomes are currently being assembled, such as the duck, the golden eagle, the medium ground finch, the budgerigar, the brown kiwi and the flycatcher.

3.2. The chicken DNA Methylation Landscape

The advent of HTS allowed not only to sequence a large number of bird genomes, but also to study their epigenomes at a whole genome scale. In 2011, Li and colleagues provided the first global DNA methylation analysis in the liver and muscle tissues of red jungle fowl and avian broiler by MeDIP-seq [74]. They observed a decrease of DNA methylation level in the promoter region, especially at the transcription start site (TSS), and an increase in the gene body regions [74]. This is consistent with the observation that average methylation patterns were the lowest in CpG islands that are localized in the promoters [75]. Slight methylation variations were observed along the gene body [74]. It was found that the methylation density in introns was higher than in untranslated transcribed regions (UTRs) and exons within the gene body [75]. The methylation profiles of proximal promoter regions of the avian broiler (liver and muscle) and of the red jungle fowl liver were negatively correlated with gene expression. These results followed the same trend as DNA methylation profile in human (for a review see [76]). Methylation level was shown to vary among tissues, the hepatic DNA of red jungle fowl being found to be less methylated than the one of the muscle [74]. This methylation variation between tissues could play a role in the regulation of gene expression during tissue differentiation [77]. Despite the density of interspersed repeats is lower in chicken (11%) than in mammals (40–50%) [67], the methylation profiles were similar with about 60% of the repeated sequences being totally methylated, which was confirmed by Hu et al. [75]. Thus, several similarities with mammals can be observed in chicken DNA methylation patterns.

The first single-base resolution chicken methylome study was published in 2015 in a study focusing on DNA methylation variation in disease resistance (Marek's disease) of two chicken lines [78], six years after the first single-base resolution Human methylome [79]. This WGBS analysis had a higher resolution and coverage than the two previous studies described above. The DNA methylation profile of lungs was performed in two chicken lines, Fayoumi and Leghorn. Li and colleagues confirmed that DNA methylation variation is likely to confer disease resistance through gene regulation [78]. Within genes, they found the highest methylation level in internal exons, which is consistent with a previous study [80]. Toll-like receptor 4 (TLR4), an immunity-related gene, was found both differentially methylated and expressed between the Fayoumi and Leghorn lines [78]. A previous study has revealed that a susceptible chicken line had a greater methylation level of the *TLR4* gene after infection [81] underlining the important role of this gene in the immune response. Thus, this study also highlighted the potential regulatory role of DNA methylation in the control of gene expression. Another recent WGBS study lead to new insights about DNA methylation variations during the lifespan of chicken. Indeed, Zhang et al. revealed variation in methylation profiles at different ages in hens breast muscle tissue [82]. The authors showed that late-laying hens showed a higher DNA methylation level along the genome compared to the juvenile hens. This increase in methylation was also found at the promoter of genes known to be involved in meat quality [82]. These findings are similar to what Laine et al. reported using a passerine bird species [9]. Indeed, the authors investigated the Great Tit methylation profiles in whole brain and blood tissue by WGBS. They found reduced methylation levels within CpG islands and at the TSS and a negative correlation between gene-body methylation and gene expression in the brain. Derks et al. [83] further investigated the Great Tit methylome and observed a higher CpG methylation level in brain compared to blood. The authors reported a link between brain TSS methylation, gene function and the activity of transposable element that were hypermethylated at CpG dinucleotides but hypomethylated at non-CpG sites. The non-CpG methylation rate was negatively correlated with transposable element activity in the brain and it was proposed that non-CpG methylation might be involved in the regulation of transposable element silencing in this tissue. The significant presence of non-CpG methylation in brain is in agreement with what was reported in mammals and chicken [78,82,83]. Altogether, these studies suggest that the methylation landscape is likely to be comparable between bird species and other species such as mammals.

3.3. The Chicken HPTM Landscape

The first bird ChIP-seq experiment was published in 2012 [84,85]. To date, only few studies reported histone modifications maps in birds [84–87]. Song's group investigated HPTM enrichment in Marek's disease infection in two lines of chicken that differ by their resistance to disease [84–86]. A genome-wide map of H3K4me3 and H3K27me3 was performed on spleen [84], thymus [85] and the Bursa of Fabricius [86], three tissues implicated in the immune response. The authors observed a sharp repartition of H3K4me3 surrounding the TSS, as previously described in mammals [61]. This was also confirmed recently in chicken erythrocytes [87]. H3K4me3 was associated with active genes, similarly to what was reported for mammalian species. The repressive mark H3K27me3 was broadly enriched at the TSS and in the body of genes [84–86]. H3K27me3 was shown to be associated with repressed genes as in other species [61,65]. Interestingly, these comparative analyses revealed differentially enriched regions between susceptible and resistant chicken lines [85,86], suggesting the existence of an "epigenetic signature" for disease susceptibility. Moreover, Song and collaborators demonstrated that the HPTM profile was modulated after infection at the loci of genes involved in the immune system response [84], corresponding to expression changes after infection [88]. For example, at the locus of *CD8 α* in spleen (for which expression was slightly or highly increased after infection in susceptible and resistant chicken lines, respectively), H3K4me3 was increased concomitantly to gene expression depending on the line [84]. In thymus, a similar trend was observed at the *CTLA-4* locus [85]. Recently, the chicken erythrocyte description of H3K27ac and H3K4me3 profile by ChIP-seq

was released [87] (unlike mammals, avian erythrocytes are nucleated). As expected, these HPTM were associated with active genes and their genome-wide distribution was similar to what has been reported for other cell types [61,89].

In conclusion, these whole genome studies in several chicken lines [84–87] all underline similar epigenetic profiles between birds and mammals. This is in agreement with the fact that epigenetic mechanisms are highly conserved between species ranging from plants to mammals [76]. They confirm that epigenetic mechanisms are likely to be a contributor in the phenotypic plasticity in bird through at least DNA methylation and HPTM changes.

4. Design of Epigenomic Studies

HTS techniques are in continuous evolution and are nowadays well documented in extensively studied model species such as human, mouse, *Drosophila* and *Arabidopsis*. Here we summarize considerations for avian epigenome investigation from sequencing considerations to the data analysis involving huge datasets. The importance of a carefully designed and executed experimental design has been addressed in previous reports [90,91]. In particular, experimental design should take into account parental history (such as rearing conditions and nutrition) and hatchery practices. Confounding factors such as flock age, transport of eggs and storage or care of animals should be minimized as much as possible, while the tissue homogeneity has to be maximized. The use of inbred strains may also provide an interesting opportunity to study epigenetic variations in a situation in which genetic variation is greatly minimized and therefore facilitates the interpretation of epigenetic changes, especially for the detection of DNA methylation based on the bisulfite-induced polymorphism [92].

4.1. Sequencing Considerations

The first study concerning DNA methylation on whole bisulfite-converted genome was published in 2008 on *Arabidopsis thaliana* at a single-CpG resolution with approximately 20-fold sequencing coverage [93,94]. Rapidly, many others studies have adapted this bisulfite method (WGBS, RRBS) in different species, particularly in human [79] and mouse [95]. The chemical bisulfite treatment involved in both techniques requires a DNA of sufficient quality (base integrity) in order to obtain an efficient conversion [96]. Choosing between RRBS and WGBS depends on the aim of the study [97]. In contrast to WGBS, the RRBS method provides a high number of reads for each studied cytosine at a reasonable cost [46]. However, RRBS induces an enrichment in regions with high CpG content that may bias the analysis toward promoters. Therefore, WGBS may be more relevant for the study of discrete methylation changes such as in gene bodies. Sequencing is a key step in the detection of epigenetic information, the reliability of analyses being not only linked to the number of replicates, but also to the sequencing depth. The sequencing depth was defined by Stein et al. [98] as “the average number of times that a particular nucleotide is represented in a collection of random raw sequences”. For WGBS, the Epigenomics Roadmap recommends from 5X to 30X of coverage [99] and about 10X of coverage for RRBS as 1–2% of the genome is sequenced using this technique [99]. To our knowledge, no specific chicken RRBS study has been published yet, therefore the pattern of MspI restriction enzyme in chicken should be carefully examined to estimate the fraction of the genome that will be included in such studies.

Since the first whole genome study of HPTM on human cells in 2007 [61], ChIP-seq became an affordable technique due to improvements in the methodology and sequencing techniques. In 2012, the ENCODE consortium published a guideline for conducting ChIP-seq experiments. Standards were defined to characterize HPTM such as the antibody quality, the control used, the number of replicates and the number of unique reads obtained after sequencing [100]. Although these recommendations were initially defined for mammalian species, they should be followed for ChIP-seq studies in birds, given the similarities between the genomes. The ENCODE consortium recommended 20 million to 40 million reads for ChIP-seq analysis according to the type of mark studied [100]. For the broad-source H3K27me3 mark, a high coverage of 40–50 million unique reads is recommended to identify regions

of interest in humans [101]. 40 millions of unique reads appeared to be sufficient for H3K27me3 in birds [65]. It is worth noting that native preparation of chromatin may be more suitable for ChIP-seq experiment in chicken muscle than the usual protocol of chromatin fixation [65].

4.2. Bioinformatic Data Analysis

The computational analysis of the data is one of the most critical, yet often underestimated element of a genome-wide studies. It has become standard for sequencing core facilities to provide basic bioinformatics services to process the generated data. However, achieving a complete, accurate and reliable analysis of epigenome data with the latest version of a genome and tools still remains challenging. The main reason is the fast pace of change in the field of bioinformatics that constantly affects the stability of any analytic pipeline. Minor format changes of the HTS data can invalidate a robust pipeline (as illustrated by the evolution of the Phred quality score across Sanger/Solexa/Illumina FASTQ formats). Biotechnological improvement results in a constantly increasing volume of data, that requires new storage formats and processing algorithms. For example, regarding sequence alignment, hitherto reference approaches like BLAST/BLAT have been shown to be inefficient to deal with billions of short reads, and dedicated mappers were designed based on new algorithmic applications [102,103]. Moreover, methodological advances added another gear to the frantic pace of bioinformatics evolution. For instance, implementing the Burrows-Wheeler Transformation in the context of read mapping induced a major shift considering the previous hash-based approaches [102,103], with direct consequences in epigenomics [60]. Indeed, popular alignment tools such as Bowtie [102], BWA [103] and SOAP2 [104] have specific alignment method and parameters (allowed mismatches) that impact the percentage of the successfully-mapped reads and ultimately the peak calling [105]. These specificities were integrated in several pipelines developed to study HTS data such as the ENCODE's WGBS pipeline [106] or the BLUEPRINT's ChIP-Seq pipeline [107]. Similar standards are expected for avian species in the context of the FAANG consortium [41].

Moreover, two factors have to be considered regarding the computational resources: computing capacity in terms of CPU and memory to run the analysis, and disk space to store data and results. Bioinformatics facilities usually provide cluster or grid computing systems to cope with the first aspect. The experimental design and the budget of a given project are often considered beforehand and the volume of raw sequencing reads is often defined by the objective (sequencing depth). Estimations are often less accurate for the total amount of data to be stored, at least during the time of the analysis. Intermediate files largely contribute to the inflation in terms of disk space. Indeed, a typical analysis pipeline or workflow is made of several steps (as shown in Figure 3 with an analytic pipeline defined for epigenetics marks). Each step generates intermediate results to be written on the disk in a precise file format before being processed by the next step. While the general trend is a drastic reduction of the data along the pipeline (from large HTS read files to relatively small end-results), initial operations like cleaning the reads or mapping on the genomic sequence produce large files. For instance, about 35 Go of data are currently obtained for the whole analysis for each WGBS sample in chicken (data not shown), necessitating adequate storage capacity. For chicken ChIP-seq about 10 Go of disk storage (data not shown) are required for the analysis of one mark (including the corresponding input) from the initial compressed fastq files (~1–2 Go) to the peak detection files (500 Ko). Intermediate results can of course be eliminated as soon as they are processed, but given that an analytic pipeline is rarely used only once in practice (e.g., when a change in the data occurs such as quality filtering or when a software is improved), it would be time- and resource-consuming to generate them again each time when required. Moreover, systematically comparing results from successive runs instead of overwriting previous ones is part of the quality control of data. Depending on the experimental assay, it is therefore advisable to secure the provision of at least 3–5 times the volume of raw data for downstream processing.

Besides, the computing capacity cluttering is further amplified when adapting or optimizing the pipeline (as often required for non-model organisms), since testing different parameters or software involves the comparison of many runs. To cope with these difficulties, needs in both human and

computational resources should not be underestimated, especially when dealing with avian species. Because most of the analytic tools have been created, tested and optimized on model organisms (human, yeast, mouse), applying a standard pipeline designed for the 22 + XY human chromosomes on the 32 + ZW chromosomes +23,870 scaffolds of the *Gallus gallus* genome (Gallus_gallus-5.0; GCA_000002315.3) can involve some tweaking.

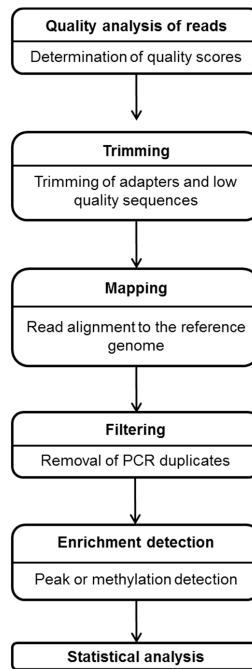


Figure 3. Example of a bioinformatics pipeline to analyze WGBS or ChIP-seq data. The quality of the sequencing reads is estimated by fastqc metrics such as the Q score [108]. Reads are usually considered of good quality when their score is above 30 (at q30 the base call accuracy is 99.9%, [109]). Optionally, reads can be trimmed to remove adapter sequences and low quality bases [110]. Reads are then mapped to the reference genome [111]. An optional filtering step can be performed by removing PCR duplicates to conserve only uniquely mapped reads before the enrichment detection. The enrichment detection step is specific to the targeted epigenetic mark. Several tools have been developed to characterize the DNA methylation across the genome (for a review see [112]) or to study HPTM depending on their repartition on the genome as HPTM were shown to exhibit sharp or broad peaks of enrichment (for a review see [113]). Finally, statistical differential analyses can be performed when two or more conditions have to be compared [112,113].

5. Conclusions and Perspectives

In this review, we aimed to summarize the knowledge gained on avian epigenomes by focusing on the most studied avian species, the chicken. The development of HTS techniques allowed the study of several epigenetic marks in chicken such as DNA methylation and histones post-translational modifications. Genome-wide characterization of these marks demonstrated similarities between mammalian and plant epigenomes. The chicken methylome and the HPTM landscape exhibited variations along the genome linked with genomic features such as the TSS of genes. Moreover, as previously reported in mammals, H3K4me3 and H3K27ac enrichments at promoters were correlated with gene expression and DNA methylation, while H3K27me3 enrichment was usually associated with repression. Epigenetic modifications and their impact on gene expression are complex phenomena that can be better understood with the sensitivity of HTS and its genome-wide DNA binding site detection. HTS-based methods were performed on different chicken breeds to investigate epigenome

variations associated with disease resistance, such as Marek's disease. DNA methylation and HPTM differences have been uncovered on genes involved in the immune response between disease-resistant and -susceptible chicken lines. In addition, it has recently been revealed that methylation profiles could vary during the lifespan of chicken in muscle and blood cells. Beyond DNA methylation and HPTM cartography, the avian community continues to improve our understanding of bird epigenomes by combining complementary epigenetic approaches, such as the mapping of open chromatin sites by ATAC-seq [114] or the analysis of genome conformation by Hi-C [115]. In the future, this should deliver a much clearer view of the mechanisms regulating genome activity related to organ function, disease or phenotypic variation.

International consortia provided guidelines to standardize the use of HTS techniques in order to detect epigenetic marks in several species that should be taken into consideration for avian studies. Bioinformatics analyses represent a key step with its own limitations and recommendations that are inherent to an epigenome study. Unexpected results or variations of the experimental protocol might also require specific skills or expertise to properly adapt the downstream analysis.

During the past decade, HTS has seen its cost and time consumption dramatically decreased, allowing the sequencing and the annotation of several bird species genomes. The necessity of a properly assembled and annotated genome could explain why only few avian epigenomes have been characterized to date. In addition to chicken studies, the investigation of other avian epigenomes, such as the Great Tit [9], should also deliver significant knowledge in the field of developmental biology.

Acknowledgments: This work was supported by the Region Centre Val de Loire, by the Department of Animal Genetics and the Department of Animal Physiology and Livestock Systems at the French National Institute of Agronomical Research (INRA). This work is funded by French ANR "ChickStress" grant (ANR-13-ADAP-0014) and by the INRA PHASE division project "STRESSEPIMARK".

Author Contributions: All authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BS-seq	Bisulfite Sequencing
ChIP-seq	Chromatin ImmunoPrecipitation followed by sequencing
ENCODE	ENCyclopedia of DNA Elements
FAANG	Functional Annotation of ANimal Genomes
Go	Giga-octet
H3K27ac	histone H3 lysine 27 acetylation
H3K27me3	histone H3 lysine 27 tri-methylation
H3k36me3	histone H3 lysine 36 tri-methylation
H3K4me3	histone H3 lysine 4 tri-methylation
H3K9me3	histone H3 lysine 9 tri-methylation
HPTM	Histone Post Translational Modifications
HTS	High-Throughput Sequencing
IHEC	International Human Epigenome Consortium
Ko	Kilo-octet
MBD-seq	Methyl Binding Domain sequencing
MeDIP-seq	METHylated DNA ImmunoPrecipitation sequencing
MethylC-seq	MethylC-sequencing
modENCODE	Model Organism ENCyclopedia Of DNA Elements
MRE-seq	Methyl-sensitive Restriction Enzyme digestion sequencing
N-ChIP-seq	Native Chromatin ImmunoPrecipitation followed by sequencing
RRBS	Reduced Representation Bisulfite Sequencing
TSS	Transcription Start Site
UTR	Untranslated Transcribed Region
WGBS	Whole Genome Bisulfite Sequencing
X-ChIP-seq	Cross-linked Chromatin ImmunoPrecipitation followed by sequencing

References

1. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753. [[CrossRef](#)] [[PubMed](#)]
2. Triantaphyllopoulos, K.A.; Ikonomopoulos, I.; Bannister, A.J. Epigenetics and inheritance of phenotype variation in livestock. *Epigenet. Chromatin* **2016**, *9*, 31. [[CrossRef](#)] [[PubMed](#)]
3. Nätt, D.; Lindqvist, N.; Stranneheim, H.; Lundeberg, J.; Torjesen, P.A.; Jensen, P. Inheritance of acquired behaviour adaptations and brain gene expression in chickens. *PLoS ONE* **2009**, *4*, e6405. [[CrossRef](#)]
4. Verhulst, E.C.; Mateman, A.C.; Zwier, M.V.; Caro, S.P.; Verhoeven, K.J.F.; van Oers, K. Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Mol. Ecol.* **2016**, *25*, 1801–1811. [[CrossRef](#)] [[PubMed](#)]
5. Kisliouk, T.; Meiri, N. A critical role for dynamic changes in histone H3 methylation at the Bdnf promoter during postnatal thermotolerance acquisition. *Eur. J. Neurosci.* **2009**, *30*, 1909–1922. [[CrossRef](#)] [[PubMed](#)]
6. Yossifoff, M.; Kisliouk, T.; Meiri, N. Dynamic changes in DNA methylation during thermal control establishment affect CREB binding to the brain-derived neurotrophic factor promoter. *Eur. J. Neurosci.* **2008**, *28*, 2267–2277. [[CrossRef](#)] [[PubMed](#)]
7. Li, S.; Zhi, L.; Liu, Y.; Shen, J.; Liu, L.; Yao, J.; Yang, X. Effect of in ovo feeding of folic acid on the folate metabolism, immune function and epigenetic modification of immune effector molecules of broiler. *Br. J. Nutr.* **2016**, *115*, 411–421. [[CrossRef](#)] [[PubMed](#)]
8. Hu, Y.; Sun, Q.; Li, X.; Wang, M.; Cai, D.; Li, X.; Zhao, R. In Ovo injection of betaine affects hepatic cholesterol metabolism through epigenetic gene regulation in newly hatched chicks. *PLoS ONE* **2015**, *10*, e0122643. [[CrossRef](#)] [[PubMed](#)]

9. Laine, V.N.; Gossman, T.I.; Schachtschneider, K.M.; Garroway, C.J.; Madsen, O.; Verhoeven, K.J.F.; De Jager, V.; Megens, H.J.; Warren, W.C.; Minx, P.; et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **2016**, *7*, 10474. [[CrossRef](#)] [[PubMed](#)]
10. Allis, C.D.; Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **2016**, *17*, 487–500. [[CrossRef](#)] [[PubMed](#)]
11. Roundtree, I.A.; Evans, M.E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017**, *169*, 1187–1200. [[CrossRef](#)] [[PubMed](#)]
12. Berger, S.L.; Kouzarides, T.; Shiekhata, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23*, 781–783. [[CrossRef](#)] [[PubMed](#)]
13. Feil, R.; Fraga, M.F. Epigenetics and the environment: Emerging patterns and implications. *Nat. Rev. Genet.* **2012**, *13*, 97–109. [[CrossRef](#)] [[PubMed](#)]
14. Egger, G.; Liang, G.; Aparicio, A.; Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, *429*, 457–463. [[CrossRef](#)] [[PubMed](#)]
15. Bock, C. Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* **2012**, *13*, 705–719. [[CrossRef](#)] [[PubMed](#)]
16. Gardiner-Garden, M.; Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **1987**, *196*, 261–282. [[CrossRef](#)]
17. Wu, T.P.; Wang, T.; Seetin, M.G.; Lai, Y.; Zhu, S.; Lin, K.; Liu, Y.; Byrum, S.D.; Mackintosh, S.G.; Zhong, M.; et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **2016**, *532*, 329–333. [[CrossRef](#)] [[PubMed](#)]
18. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **1986**, *321*, 209–213. [[CrossRef](#)] [[PubMed](#)]
19. Keshet, I.; Lieman-Hurwitz, J.; Cedar, H. DNA methylation affects the formation of active chromatin. *Cell* **1986**, *44*, 535–543. [[CrossRef](#)]
20. Razin, A.; Cedar, H. DNA methylation and gene expression. *Microbiol. Rev.* **1991**, *55*, 451–458. [[PubMed](#)]
21. Bird, A. The essentials of DNA methylation. *Cell* **1992**, *70*, 5–8. [[CrossRef](#)]
22. Jones, P.A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [[CrossRef](#)] [[PubMed](#)]
23. Busche, S.; Shao, X.; Caron, M.; Kwan, T.; Allum, F.; Cheung, W.A.; Susan Westfall, B.G.; Simon, M.-M. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. *Genome Biol.* **2015**, *16*, 290. [[CrossRef](#)] [[PubMed](#)]
24. Wolff, G.L.; Kodell, R.L.; Moore, S.R.; Cooney, C.A. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J.* **1998**, *12*, 949–957. [[PubMed](#)]
25. Kornberg, R.D.; Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **1999**, *98*, 285–294. [[CrossRef](#)]
26. Strahl, B.D.; Allis, C.D. The language of covalent histone modifications. *Nature* **2000**, *403*, 41–45. [[CrossRef](#)] [[PubMed](#)]
27. Tessarz, P.; Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 703–708. [[CrossRef](#)] [[PubMed](#)]
28. Bannister, A.J.; Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381–395. [[CrossRef](#)] [[PubMed](#)]
29. Kouzarides, T. Chromatin modifications and their function. *Cell* **2007**, *128*, 693–705. [[CrossRef](#)] [[PubMed](#)]
30. Eberharter, A.; Becker, P.B. Histone acetylation: A switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep.* **2002**, *3*, 224–229. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, T.; Cooper, S.; Brockdorff, N. The interplay of histone modifications—Writers that read. *EMBO Rep.* **2015**, *16*, 1467–1481. [[CrossRef](#)] [[PubMed](#)]
32. Lämke, J.; Brzezinka, K.; Altmann, S.; Bäurle, I. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J.* **2016**, *35*, 162–175. [[CrossRef](#)] [[PubMed](#)]
33. Coustham, V.; Li, P.; Strange, A.; Lister, C.; Song, J.; Dean, C. Quantitative modulation of polycomb silencing underlies natural variation in vernalization. *Science* **2012**, *337*, 584–587. [[CrossRef](#)] [[PubMed](#)]
34. Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; Ziller, M.J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. [[CrossRef](#)] [[PubMed](#)]

35. Baker, M. Making sense of chromatin states. *Nat. Methods* **2011**, *8*, 717–722. [[CrossRef](#)] [[PubMed](#)]
36. Ernst, J.; Kheradpour, P.; Mikkelsen, T.S.; Shores, N.; Ward, L.D.; Epstein, C.B.; Zhang, X.; Wang, L.; Issner, R.; Coyne, M.; et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **2011**, *473*, 43–49. [[CrossRef](#)] [[PubMed](#)]
37. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)]
38. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048. [[CrossRef](#)] [[PubMed](#)]
39. Bae, J.-B. Perspectives of International Human Epigenome Consortium. *Genom. Inform.* **2013**, *11*, 7. [[CrossRef](#)] [[PubMed](#)]
40. Celniker, S.E.; Dillon, L.A.L.; Gerstein, M.B.; Gunsalus, K.C.; Henikoff, S.; Karpen, G.H.; Kellis, M.; Lai, E.C.; Lieb, J.D.; MacAlpine, D.M.; et al. Unlocking the secrets of the genome. *Nature* **2009**, *459*, 927–930. [[CrossRef](#)] [[PubMed](#)]
41. Andersson, L.; Archibald, A.L.; Bottema, C.D.; Brauning, R.; Burgess, S.C.; Burt, D.W.; Casas, E.; Cheng, H.H.; Clarke, L.; Couldrey, C.; et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* **2015**, *16*, 57. [[CrossRef](#)] [[PubMed](#)]
42. Kurdyukov, S.; Bullock, M. DNA Methylation Analysis: Choosing the Right Method. *Biology* **2016**, *5*, 3. [[CrossRef](#)] [[PubMed](#)]
43. Frommer, M.; McDonald, L.E.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831. [[CrossRef](#)] [[PubMed](#)]
44. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [[CrossRef](#)] [[PubMed](#)]
45. Urich, M.A.; Nery, J.R.; Lister, R.; Schmitz, R.J.; Ecker, J.R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **2015**, *10*, 475–483. [[CrossRef](#)] [[PubMed](#)]
46. Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005**, *33*, 5868–5877. [[CrossRef](#)] [[PubMed](#)]
47. Gu, H.; Smith, Z.D.; Bock, C.; Boyle, P.; Gnirke, A.; Meissner, A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* **2011**, *6*, 468–481. [[CrossRef](#)] [[PubMed](#)]
48. Soto, J.; Rodriguez-Antolin, C.; Vallespín, E.; de Castro Carpeño, J.; Ibanez de Caceres, I. The impact of next-generation sequencing on the DNA methylation-based translational cancer research. *Transl. Res.* **2016**, *169*, 1–18. [[CrossRef](#)] [[PubMed](#)]
49. Harris, R.A.; Wang, T.; Coarfa, C.; Nagarajan, R.P.; Hong, C.; Downey, S.L.; Johnson, B.E.; Fouse, S.D.; Delaney, A.; Zhao, Y.; et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* **2010**, *28*, 1097–1105. [[CrossRef](#)] [[PubMed](#)]
50. Lan, X.; Adams, C.; Landers, M.; Dudas, M.; Krüssinger, D.; Marnellos, G.; Bonneville, R.; Xu, M.; Wang, J.; Huang, T.H.M.; et al. High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. *PLoS ONE* **2011**, *6*, e22226. [[CrossRef](#)] [[PubMed](#)]
51. Dolinoy, D.C.; Huang, D.; Jirtle, R.L. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13056–13061. [[CrossRef](#)] [[PubMed](#)]
52. Pértille, F.; Guerrero-Bosagna, C.; Da Silva, V.H.; Boschiero, C.; da Silva Nunes, J.D.R.; Ledur, M.C.; Jensen, P.; Coutinho, L.L. High-throughput and Cost-effective Chicken Genotyping Using Next-Generation Sequencing. *Sci. Rep.* **2016**, *6*, 26929. [[CrossRef](#)] [[PubMed](#)]
53. Pértille, F.; Brantsæter, M.; Nordgreen, J.; Coutinho, L.L.; Janczak, A.M.; Jensen, P.; Guerrero-Bosagna, C. DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. *J. Exp. Biol.* **2017**, *220*, 3579–3587. [[CrossRef](#)] [[PubMed](#)]

54. Kriaucionis, S.; Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009**, *324*, 929–930. [[CrossRef](#)] [[PubMed](#)]
55. Tahiliani, M.; Koh, K.P.; Shen, Y.; Pastor, W.A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L.M.; Liu, D.R.; Aravind, L.; et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935. [[CrossRef](#)] [[PubMed](#)]
56. Ponnaluri, V.K.C.; Ehrlich, K.C.; Zhang, G.; Lacey, M.; Johnston, D.; Pradhan, S.; Ehrlich, M. Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression. *Epigenetics* **2017**, *12*, 123–138. [[CrossRef](#)] [[PubMed](#)]
57. Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **2010**, *5*, e15367. [[CrossRef](#)] [[PubMed](#)]
58. Booth, M.J.; Ost, T.W.; Beraldi, D.; Bell, N.M.; Branco, M.R.; Reik, W.; Balasubramanian, S. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* **2013**, *8*, 1841–1851. [[CrossRef](#)] [[PubMed](#)]
59. Hahn, M.A.; Li, A.X.; Wu, X.; Pfeifer, G.P. Single base resolution analysis of 5-methylcytosine and 5-hydroxymethylcytosine by RRBS and TAB-RRBS. *Methods Mol. Biol.* **2015**, *1238*, 273–287. [[CrossRef](#)] [[PubMed](#)]
60. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. [[CrossRef](#)] [[PubMed](#)]
61. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007**, *129*, 823–837. [[CrossRef](#)] [[PubMed](#)]
62. Johnson, D.S.; Mortazavi, A.; Myers, R.M.; Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **2007**, *316*, 1497–1502. [[CrossRef](#)] [[PubMed](#)]
63. Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.; et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **2007**, *4*, 651–657. [[CrossRef](#)] [[PubMed](#)]
64. Edwards, J.R.; O'Donnell, A.H.; Rollins, R.A.; Peckham, H.E.; Lee, C.; Milekic, M.H.; Chanrion, B.; Fu, Y.; Su, T.; Hibshoosh, H.; et al. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res.* **2010**, *20*, 972–980. [[CrossRef](#)] [[PubMed](#)]
65. David, S.A.; Piégu, B.; Hennequet-Antier, C.; Pannetier, M.; Aguirre-Lavin, T.; Crochet, S.; Bordeau, T.; Couroussé, N.; Brionne, A.; Bigot, Y.; et al. An Assessment of Fixed and Native Chromatin Preparation Methods to Study Histone Post-Translational Modifications at a Whole Genome Scale in Skeletal Muscle Tissue. *Biol. Proced. Online* **2017**, *19*, 10. [[CrossRef](#)] [[PubMed](#)]
66. Rhee, H.S.; Pugh, B.F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **2011**, *147*, 1408–1419. [[CrossRef](#)] [[PubMed](#)]
67. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **2004**, *432*, 695–716. [[CrossRef](#)]
68. McQueen, H.A.; Fantes, J.; Cross, S.H.; Clark, V.H.; Archibald, A.L.; Bird, A.P. CpG islands of chicken are concentrated on microchromosomes. *Nat. Genet.* **1996**, *12*, 321–324. [[CrossRef](#)] [[PubMed](#)]
69. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [[CrossRef](#)] [[PubMed](#)]
70. Warren, W.C.; Hillier, L.W.; Tomlinson, C.; Minx, P.; Kremitzki, M.; Graves, T.; Markovic, C.; Bouk, N.; Pruitt, K.D.; Thibaud-Nissen, F.; et al. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 Genes Genomes Genet.* **2017**, *7*, 109–117. [[CrossRef](#)] [[PubMed](#)]
71. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)] [[PubMed](#)]
72. Dalloul, R.A.; Long, J.A.; Zimin, A.V.; Aslam, L.; Beal, K.; Blomberg, L.A.; Bouffard, P.; Burt, D.W.; Crasta, O.; Crooijmans, R.P.; et al. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol.* **2010**, *8*, e1000475. [[CrossRef](#)] [[PubMed](#)]
73. NCBI. Eukaryotic Genomes Annotated at NCBI. Available online: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/ (accessed on 6 July 2017).

74. Li, Q.; Li, N.; Hu, X.; Li, J.; Du, Z.; Chen, L.; Yin, G.; Duan, J.; Zhang, H.; Zhao, Y.; et al. Genome-wide mapping of DNA methylation in chicken. *PLoS ONE* **2011**, *6*, e19428. [[CrossRef](#)] [[PubMed](#)]
75. Hu, Y.; Xu, H.; Li, Z.; Zheng, X.; Jia, X.; Nie, Q.; Zhang, X. Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers. *PLoS ONE* **2013**, *8*, e56411. [[CrossRef](#)] [[PubMed](#)]
76. Suzuki, M.M.; Bird, A. DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet.* **2008**, *9*, 465–476. [[CrossRef](#)] [[PubMed](#)]
77. Xu, Q.; Zhang, Y.; Sun, D.; Wang, Y.; Yu, Y. Analysis on DNA methylation of various tissues in chicken. *Anim. Biotechnol.* **2007**, *18*, 231–241. [[CrossRef](#)] [[PubMed](#)]
78. Li, J.; Li, R.; Wang, Y.; Hu, X.; Zhao, Y.; Li, L.; Feng, C.; Gu, X.; Liang, F.; Lamont, S.J.; et al. Genome-wide DNA methylome variation in two genetically distinct chicken lines using MethylC-seq. *BMC Genom.* **2015**, *16*, 851. [[CrossRef](#)] [[PubMed](#)]
79. Lister, R.; Pelizzola, M.; Dowen, R.H.; Hawkins, R.D.; Hon, G.; Tonti-Filippini, J.; Nery, J.R.; Lee, L.; Ye, Z.; Ngo, Q.M.; et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **2009**, *462*, 315–322. [[CrossRef](#)] [[PubMed](#)]
80. Tian, F.; Zhan, F.; VanderKraats, N.D.; Hiken, J.F.; Edwards, J.R.; Zhang, H.; Zhao, K.; Song, J. DNMT gene expression and methylome in Marek's disease resistant and susceptible chickens prior to and following infection by MDV. *Epigenetics* **2013**, *8*, 431–444. [[CrossRef](#)] [[PubMed](#)]
81. Gou, Z.; Liu, R.; Zhao, G.; Zheng, M.; Li, P.; Wang, H.; Zhu, Y.; Chen, J.; Wen, J. Epigenetic modification of TLRs in leukocytes is associated with increased susceptibility to *Salmonella enteritidis* in chickens. *PLoS ONE* **2012**, *7*, e33627. [[CrossRef](#)] [[PubMed](#)]
82. Zhang, M.; Yan, F.B.; Li, F.; Jiang, K.R.; Li, D.H.; Han, R.L.; Li, Z.J.; Jiang, R.R.; Liu, X.J.; Kang, X.T.; et al. Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Sci. Rep.* **2017**, *7*, 45564. [[CrossRef](#)] [[PubMed](#)]
83. Derks, M.F.L.; Schachtschneider, K.M.; Madsen, O.; Schijlen, E.; Verhoeven, K.J.F.; van Oers, K. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genom.* **2016**, *17*, 332. [[CrossRef](#)] [[PubMed](#)]
84. Luo, J.; Mitra, A.; Tian, F.; Chang, S.; Zhang, H.; Cui, K.; Yu, Y.; Zhao, K.; Song, J. Histone methylation analysis and pathway predictions in chickens after MDV infection. *PLoS ONE* **2012**, *7*, e41849. [[CrossRef](#)] [[PubMed](#)]
85. Mitra, A.; Luo, J.; Zhang, H.; Cui, K.; Zhao, K.; Song, J. Marek's disease virus infection induces widespread differential chromatin marks in inbred chicken lines. *BMC Genom.* **2012**, *13*, 557. [[CrossRef](#)] [[PubMed](#)]
86. Mitra, A.; Luo, J.; He, Y.; Gu, Y.; Zhang, H.; Zhao, K.; Cui, K.; Song, J. Histone modifications induced by MDV infection at early cytolytic and latency phases. *BMC Genom.* **2015**, *16*, 311. [[CrossRef](#)] [[PubMed](#)]
87. Jahan, S.; Xu, W.; He, S.; Gonzalez, C.; Delcuve, G.P.; Davie, J.R. The chicken erythrocyte epigenome. *Epigenet. Chromatin* **2016**, *9*, 19. [[CrossRef](#)] [[PubMed](#)]
88. Yu, Y.; Luo, J.; Mitra, A.; Chang, S.; Tian, F.; Zhang, H.; Yuan, P.; Zhou, H.; Song, J. Temporal transcriptome changes induced by MDV in Marek's disease-resistant and -susceptible inbred chickens. *BMC Genom.* **2011**, *12*, 501. [[CrossRef](#)] [[PubMed](#)]
89. Malik, A.N.; Vierbuchen, T.; Hemberg, M.; Rubin, A.A.; Ling, E.; Couch, C.H.; Stroud, H.; Spiegel, I.; Farh, K.K.H.; Harmin, D.A.; et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **2014**, *17*, 1330–1339. [[CrossRef](#)] [[PubMed](#)]
90. Boleli, I.; Morita, V.; Matos, J., Jr.; Thimotheo, M.; Almeida, V. Poultry Egg Incubation: Integrating and Optimizing Production Efficiency. *Rev. Bras. Ciência Avícola* **2016**, *18*. [[CrossRef](#)]
91. Rigai, G.; Balzergue, S.; Brunaud, V.; Blondet, E.; Rau, A.; Rogier, O.; Caius, J.; Maugis-Rabusseau, C.; Soubigou-Taconnat, L.; Aubourg, S.; et al. Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief. Bioinform.* **2016**. [[CrossRef](#)] [[PubMed](#)]
92. Daxinger, L.; Whitelaw, E. Transgenerational epigenetic inheritance: More questions than answers. *Genome Res.* **2010**, *20*, 1623–1628. [[CrossRef](#)] [[PubMed](#)]
93. Cokus, S.J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C.D.; Pradhan, S.; Nelson, S.F.; Pellegrini, M.; Jacobsen, S.E. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **2008**, *452*, 215–219. [[CrossRef](#)] [[PubMed](#)]

94. Lister, R.; O'Malley, R.C.; Tonti-Filippini, J.; Gregory, B.D.; Berry, C.C.; Millar, A.H.; Ecker, J.R. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **2008**, *133*, 523–536. [[CrossRef](#)] [[PubMed](#)]
95. Smith, Z.D.; Gu, H.; Bock, C.; Gnirke, A.; Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **2009**, *48*, 226–232. [[CrossRef](#)] [[PubMed](#)]
96. Ziller, M.J.; Hansen, K.D.; Meissner, A.; Arye, M.J. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat. Methods* **2015**, *12*, 230–232. [[CrossRef](#)] [[PubMed](#)]
97. Yong, W.S.; Hsu, F.M.; Chen, P.Y. Profiling genome-wide DNA methylation. *Epigenet. Chromatin* **2016**, *9*, 26. [[CrossRef](#)] [[PubMed](#)]
98. Stein, L. Genome annotation: From sequence to biology. *Nat. Rev. Genet.* **2001**, *2*, 493–503. [[CrossRef](#)] [[PubMed](#)]
99. Doherty, R.; Couldrey, C. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: A technical assessment. *Front. Genet.* **2014**, *5*, 126. [[CrossRef](#)] [[PubMed](#)]
100. Landt, S.G.; Marinov, G.K.; Kundaje, A.; Kheradpour, P.; Pauli, F.; Batzoglou, S.; Bernstein, B.E.; Bickel, P.; Brown, J.B.; Cayting, P.; et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **2012**, *22*, 1813–1831. [[CrossRef](#)] [[PubMed](#)]
101. Jung, Y.L.; Luquette, L.J.; Ho, J.W.K.; Ferrari, F.; Tolstorukov, M.; Minoda, A.; Issner, R.; Epstein, C.B.; Karpen, G.H.; Kuroda, M.I.; et al. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res.* **2014**, *42*, e74. [[CrossRef](#)] [[PubMed](#)]
102. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)] [[PubMed](#)]
103. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
104. Li, R.; Yu, C.; Li, Y.; Lam, T.W.; Yiu, S.M.; Kristiansen, K.; Wang, J. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **2009**, *25*, 1966–1967. [[CrossRef](#)] [[PubMed](#)]
105. Han, Y.; He, X. Integrating Epigenomics into the Understanding of Biomedical Insight. *Bioinform. Biol. Insights* **2016**, *10*, 267–289. [[CrossRef](#)] [[PubMed](#)]
106. The ENCODE Project Consortium. Whole-Genome Bisulfite Sequencing Data Standards and Processing Pipeline. Available online: <https://www.encodeproject.org/wgbs/> (accessed on 6 July 2017).
107. Epigenome Blueprint. ChIP-Seq Analysis Pipeline. Available online: http://dcc.blueprint-epigenome.eu/#/md/chip_seq_grch37 (accessed on 6 July 2017).
108. Sun, Z.; Cunningham, J.; Slager, S.; Kocher, J.P. Base resolution methylome profiling: Considerations in platform selection, data preprocessing and analysis. *Epigenomics* **2015**, *7*, 813–828. [[CrossRef](#)] [[PubMed](#)]
109. Ewing, B.; Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **1998**, *8*, 186–194. [[CrossRef](#)] [[PubMed](#)]
110. Del Fabbro, C.; Scalabrin, S.; Morgante, M.; Giorgi, F.M. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE* **2013**, *8*, e85024. [[CrossRef](#)] [[PubMed](#)]
111. Schbath, S.; Martin, V.; Zytynicki, M.; Fayolle, J.; Loux, V.; Gibrat, J.F. Mapping reads on a genomic sequence: An algorithmic overview and a practical comparative analysis. *J. Comput. Biol.* **2012**, *19*, 796–813. [[CrossRef](#)] [[PubMed](#)]
112. Yu, X.; Sun, S. Comparing five statistical methods of differential methylation identification using bisulfite sequencing data. *Stat. Appl. Genet. Mol. Biol.* **2016**, *15*, 173–191. [[CrossRef](#)] [[PubMed](#)]
113. Steinhauser, S.; Kurzawa, N.; Eils, R.; Herrmann, C. A comprehensive comparison of tools for differential ChIP-seq analysis. *Brief. Bioinform.* **2016**, *17*, 953–966. [[CrossRef](#)] [[PubMed](#)]
114. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **2015**, *109*, 21–29. [[CrossRef](#)] [[PubMed](#)]
115. Belton, J.M.; McCord, R.P.; Gibcus, J.H.; Naumova, N.; Zhan, Y.; Dekker, J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **2012**, *58*, 268–276. [[CrossRef](#)] [[PubMed](#)]



Tableau 1.2 – Informations sur l’assemblage des versions successives de *Gallus gallus*. Le nom de l’échantillon comprend l’espèce (RJF, *Red Jungle Fowl*), le numéro de l’isolat (256) et provient de la lignée consanguine UCD 001. "NA" signifie que l’information est manquante. La taille du génome est donnée en gigabase (Gb) tandis que celle des contigs et *scaffolds* est le mégabase (Mb).

Nom de l’assemblage	Gallus_gallus-2.1	Gallus_gallus-2.1	Gallus_gallus-4.0	Gallus_gallus-5.0	GRCg6	GRCg6a
Nom de l’échantillon	RJF #256 inbred UCD001	RJF #256 inbred UCD001	RJF #256 inbred UCD001	RJF #256 inbred UCD001	RJF #256 inbred UCD001	RJF #256 inbred UCD001
Sexe	Femelle	Femelle	Femelle	Femelle	Femelle	Femelle
Date	01/11/2006	01/11/2006	22/11/2011	16/12/2015	02/02/2018	27/03/2018
Niveau d’assemblage	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome
Accession GenBank	GCA_000002315.1	GCA_000002315.1	GCA_000002315.2	GCA_000002315.3	GCA_000002315.4	GCA_000002315.5
Accession RefSeq	GCF_000002315.1	GCF_000002315.2	GCF_000002315.3	GCF_000002315.4	NA	GCF_000002315.5
Couverture du génome	6,6X	6,6X	12X	70X	82X	82X
Technologie de séquençage	Sanger, 454	Sanger, 454	Sanger, 454, Illumina	Sanger, 454, Illumina, PacBio	PacBio	PacBio
Taille du génome	1,10 Gb	1,10 Gb	1,05 Gb	1,23 Gb	1,14 Gb	1,07 Gb
Nombre de scaffolds	17506	17507	16847	23870	528	525
Taille des scaffolds (N50)	11,06 Mb	11,06 Mb	12,88 Mb	6,38 Mb	23,06 Mb	20,79 Mb
Nombre de contigs	78534	78535	27041	24693	1602	1403
Taille des contigs (N50)	0,05 Mb	0,05 Mb	0,28 Mb	2,89 Mb	16,74 Mb	17,66 Mb

1.3 Données haut-débit et outils bioinformatiques

Grâce aux avancées technologiques, l'étude de l'épigénome s'est développée dans de nombreuses espèces. Elle comprend notamment le séquençage nouvelle génération (NGS, *Next-Generation Sequencing*) à l'échelle du génome et s'appuie sur des analyses computationnelles [150]. Ces techniques permettent notamment de caractériser plus finement la méthylation de l'ADN et ainsi de mieux comprendre la régulation de l'expression des gènes. Depuis quelques années, l'avènement des techniques de séquençage haut-débit (NGS) a en outre permis de séquencer une plus grande diversité de génomes, notamment chez les oiseaux, et plus particulièrement chez la poule.

1.3.1 Amélioration de l'assemblage du génome de la poule grâce aux avancées technologiques

La première version du génome de la poule a été publiée en 2004 [151]. Il s'agit de la première espèce de rente à avoir été séquencée. Les séquences de ce génome ont été réalisées à l'époque grâce à la technique de Sanger [152]. Il s'agit d'une des deux premières techniques dont le but est de séquencer l'ADN. Contrairement à la technique de Maxam-Gilbert qui utilise des produits chimiques pour fragmenter l'ADN, Frederick Sanger propose l'idée qu'en copiant les brins d'ADN et en surveillant quels nucléotides sont ajoutés, la séquence de l'ADN peut être retrouvée. Cette technique a révolutionné la biologie en permettant d'étudier, à l'échelle des génomes, de nombreux organismes, des bactéries jusqu'à l'Homme. Des années plus tard, la deuxième vague d'avancées technologiques est appelée "nouvelle génération" (*next-generation*), ou "deuxième génération", à la suite de la technique de Sanger qualifiée de première génération. L'amélioration des technologies permet de générer de plus en plus de données en de moins en moins de temps.

Pour combler les lacunes et les duplications erronées connues dans les premières versions du génome de *Gallus gallus* (Gallus_gallus-2.1), l'ADN provenant de la même poule (RJF #256 inbred UCD001) a été re-séquencé avec les technologies Roche 454 (www.roche.com) et Illumina (www.illumina.com) pour générer une nouvelle version (Gallus_gallus-4.0). La combinaison de 454 et Illumina permet de séquencer plus vite et moins cher qu'avec du Sanger pour obtenir des génomes de bonne qualité (taux d'erreur faible et couverture uniforme). Ainsi, la troisième version de *Gallus gallus* (Gallus_gallus-4.0) est constituée d'un nouvel assemblage *de novo* provenant des lectures de Sanger précédemment générées et des nouvelles séquences 454 et Illumina. Les dernières versions de *Gallus gallus* (Gallus_gallus-5.0, GRCg6, GRCg6a) ont été enrichies par la technologie PacBio (*Pacific Biosciences RSII*, www.pacb.com), ce qui a permis d'assigner des séquences à trois chromosomes supplémentaires et d'améliorer la longueur des contigs [153]. Cette technologie de séquençage en temps réel sur molécule unique fait partie de la troisième vague d'avancées technologiques, *next-next-generation*, ou "troisième génération". Comparée aux technologies précédentes, cette troisième génération de séquenceurs haut-débit produit des séquences plus longues [154]. Il est donc intéressant de combiner les deuxièmes générations pour l'obtention de lectures courtes permettant une grande profondeur et peu d'erreurs de séquençage, et les troisièmes générations dont les lectures longues permettent de disposer de longs fragments sans assemblage préalable, afin de tirer parti de toutes leurs caractéristiques, complémentaires.

Le caryotype de la poule est constitué de 38 paires d'autosomes dont 29 sont composées de très petits chromosomes (≤ 20 Mb), de taille uniforme, indiscernables en métaphase. Ils sont appelés les microchromosomes, à la différence des 9 autres autosomes plus grands (macrochromosomes). Comme décrit précédemment (cf partie 1.2.5 page 39), les chromosomes sexuels sont ZZ pour les mâles et ZW pour les femelles chez la poule. Les premières versions des chromosomes sexuels ont été séquencées à hauteur de 3,3X en raison de l'état hémizygote de l'échantillon femelle utilisé [151]. Au fur et à mesure de l'évolution des technologies, l'assemblage du génome de la poule a été amélioré (Tableau 1.2). Pour le même échantillon, la couverture globale du génome augmente : de 6,6X à 82X. La dernière version est nettement améliorée avec une augmentation de la taille des contigs et des *scaffolds* générant un assemblage continu du génome sans interruption de séquences. Même si elle semble très complète (peu de contigs et de *scaffolds*), des régions sont toujours manquantes dans le génome de la poule. En effet, seulement 33 autosomes sont assemblés. Ce nombre important de séquences non assignées à des chromosomes nécessite des ajustements dans l'analyse bioinformatique. En effet, pour optimiser l'analyse bioinformatique, il est fréquent d'exécuter les analyses en parallèle par chromosome. La multitude de *scaffolds* (par exemple, 23870 sur la version 5) ne permet pas d'appliquer simplement ce type d'analyses. De plus, plus les quantités de données sont importantes (grand nombre de contigs et *scaffolds*), plus l'analyse sera longue et

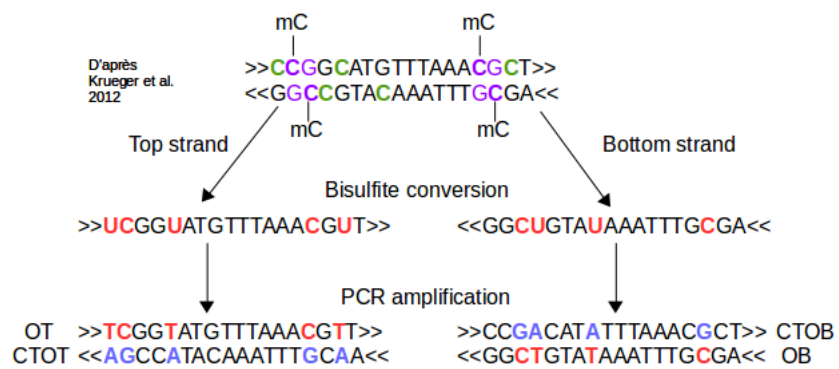


FIGURE 1.6 – Schéma du principe de la conversion au bisulfite de sodium. Le bisulfite de sodium est un agent chimique qui convertit les C non méthylés en T sur les brins de l'ADN génomique (figure issue de [155]).

coûteuse en terme de ressources informatiques.

1.3.2 Données haut-débit du méthylome

Historiquement, la détection de la méthylation de l'ADN repose sur trois principes [156]. Premièrement, elle peut être détectée directement grâce à un site de restriction enzymatique au niveau duquel l'ADN est coupé ou non par des enzymes reconnaissant la même séquence, qui sont sensibles ou non à la méthylation de l'ADN. Par exemple, pour le même site de reconnaissance sur le génome, l'enzyme MspI n'est pas sensible à la méthylation de l'ADN et donc coupe l'ADN dès qu'elle reconnaît le site (CC*GG ou CCGG), tandis que l'enzyme HpaII, sensible à la méthylation, coupe seulement si le site est non méthylé (CCGG uniquement). Deuxièmement, l'ADN méthylé peut être révélé par immunoprécipitation avec des anticorps anti-5mC voire des anticorps dirigés contre des protéines reconnaissant spécifiquement l'ADN méthylé. Et enfin, un traitement chimique de l'ADN permet de fixer l'état de méthylation, en modifiant les cytosines non méthylées en thymines. Ces méthodes peuvent être complétées avec du séquençage haut-débit, et même se combiner entre elles.

Seule la méthylation de l'ADN détectée par traitement chimique sera traitée ici.

1.3.2.1 Conversion au bisulfite

Les motifs de méthylation détectés sur tout le génome constituent le méthylome. Décrire le méthylome à la base près c'est mesurer sur chaque dinucléotide CG le taux de méthylation, qui se définit comme le pourcentage de C méthylés. La technique NGS la plus utilisée pour décrire le méthylome est appelée BS-seq (*Bisulfite-sequencing*). Il s'agit d'associer un traitement chimique au bisulfite de sodium de l'ADN suivi d'un séquençage haut-débit. Elle permet de distinguer les cytosines méthylées des autres sur les deux brins d'ADN [155]. En effet, un C méthylé n'est pas affecté par le traitement et reste donc un C alors que un C non méthylé est lu comme un T (Figure 1.6). Le traitement bisulfite des deux brins de l'ADN génomique (brin sens ou *Original Top* OT et brin anti-sens *Original Bottom* OB) convertit en uracile les cytosines non méthylées (déamination). Lors de l'amplification par PCR, les uraciles sont remplacées par des thymines. Par complémentarité, le brin synthétisé (brin sens ou *Complementary to Original Top* CTOT et brin anti-sens ou *Complementary to Original Bottom* CTOB) a donc des adénines au lieu de guanines. Une fois l'amplification réalisée, le séquençage est effectué soit en *paired-end*, c'est-à-dire séquençage de deux extrémités de fragments courts, soit en *single-end*, une seule extrémité séquencée. Les séquences obtenues ont un pourcentage de thymines important correspondant aux brins originaux OT et OB (pourcentage de cytosines très faible) et un pourcentage d'adénines élevé dans les séquences correspondant aux brins CTOT et CTOB (pourcentage de guanines très faible). Cela permet de révéler l'état de méthylation d'origine en comptant le nombre de C et de T dans les lectures traitées au bisulfite de sodium pour chaque base C de l'ADN du génome. La précision de la mesure de ce taux dépend fortement de la profondeur de séquençage (nombre moyen de lectures en une position). Une plus grande profondeur de séquençage permet une meilleure estimation de l'état de méthylation.

1.3.2.2 Séquençage haut-débit

Le génome entier de plusieurs organismes peut être analysé par séquençage haut-débit après conversion au bisulfite de sodium par WGBS (*Whole Genome Bisulfite Sequencing* ou MethylC-seq [157]). Cette technique est décrite dans notre revue [158].

Il est également possible de ne séquencer qu'une fraction du génome, pour moins cher qu'un séquençage tout-génome, ce qui permet d'analyser plus d'individus pour un coût global équivalent. Pour cela, des enzymes de restriction peuvent être utilisées pour cibler les fragments génomiques souhaités. En méthylation, la technique RRBS (*Reduced Representation Bisulfite Sequencing*) est une technique qui associe une digestion enzymatique, en général par MspI (site CCGG), et une sélection d'une fraction du génome en fonction de la taille des fragments digérés. Elle permet de détecter les cytosines dans les régions du génome enrichies en CG et d'obtenir, en général, environ 2% du génome [159]. À la différence de la technique WGBS, pour un même coût de séquençage, moins de régions sont détectées mais avec une profondeur de séquençage souvent plus importante. Généralement les fragments d'ADN entre 40 et 220 pb sont représentatifs de la majorité des séquences des promoteurs et des îlots CpG [160].

Cette technique a été utilisée dans le cadre de ma thèse et est développée dans la partie 2.2.1 page 75.

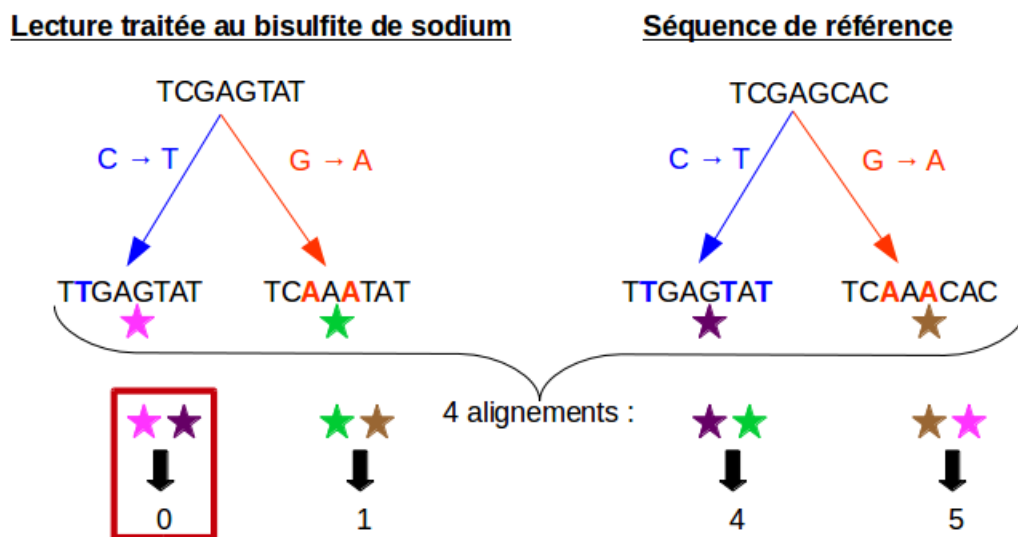


FIGURE 1.7 – Schéma de l'alignement d'une lecture traitée au bisulfite. La lecture "bisulfitée" et la séquence du génome de référence sont converties en C → T (flèche bleue) et en G → A (flèche rouge). Les 4 alignements possibles de la lecture convertie sur la référence convertie sont testés. Le nombre de bases qui ne sont pas communes à la même position entre la lecture et la référence (*mismatch*) est donné par la flèche noire. Le meilleur alignement est celui dont le nombre de *mismatches* est le plus petit, soit l'alignement dans l'encadré rouge avec 0 *mismatches* (figure issue de [161]).

1.3.3 Outils bioinformatiques d'analyses du méthylome

Le paragraphe qui suit est principalement axé sur la technique WGBS. Si des spécificités RRBS sont à souligner, cela sera mentionné.

Le développement des technologies NGS et leur coût de moins en moins élevé font de l'analyse bioinformatique une étape incontournable dans la caractérisation des épigénomes [158] : à l'issue du séquençage, des millions de séquences (ou lectures, ou *reads*) sont obtenues par échantillon. Cependant, une analyse complète et précise des données épigénomiques avec les bons outils reste un défi : les versions du génome et leur annotation sont renouvelées plus ou moins régulièrement, et l'avancée des technologies NGS demande un effort constant de mises à jour et/ou de création dans la mise en place d'outils et de pipelines bioinformatiques. Un pipeline d'analyses est un enchaînement de plusieurs étapes spécifiques aux types de données étudiées.

1.3.3.1 Pipeline d'analyses

Le pipeline d'analyses que nous avons construit pour analyser des données de méthylation issues de séquençage haut débit comprend plusieurs étapes qui sont décrites ci-dessous (cf revue [162]). Pour chaque étape d'analyses, sont introduits le but et l'outil utilisé dans le pipeline que nous avons développé.

Nettoyage des données. Tout d'abord, les lectures obtenues du séquenceur doivent être nettoyées, le but étant d'obtenir une qualité de séquence indispensable pour caractériser le phénomène observé. Les séquences d'adaptateurs qui ont servi dans le protocole expérimental (cf partie 2.2.1 page 75) et qui ne sont donc pas informatives pour la description du méthylome sont supprimées. Le logiciel Trim Galore automatise cette suppression [155] et est basé sur un outil appelé Cutadapt [163]. Le programme recherche à l'extrémité des lectures des alignements qui correspondent à la séquence de l'adaptateur et supprime les bases correspondantes. Un minimum de correspondance entre les lectures et la séquence de l'adaptateur peut être fixé (au moins un nucléotide doit correspondre par défaut). De plus, ce logiciel supporte beaucoup de formats de fichiers qui sont des standards en biologie comme le FASTA et le FASTQ. Ce sont des formats de fichiers texte qui stockent des séquences biologiques et, pour les fichiers FASTQ, la qualité des bases lues après le séquençage. Un format FASTA est constitué de deux types de lignes : l'identifiant de la lecture et sa séquence. Le format FASTQ reprend le FASTA et y associe des scores de qualité, ce qui le rend standard *de facto* pour le stockage des sorties de séquenceurs à haut débit. Afin de vérifier les résultats obtenus après avoir enlevé les séquences d'adaptateurs, l'outil FastQC présente les données sous forme de graphiques et de tableaux pour évaluer rapidement les profils des lectures [164]. Ainsi, les lectures sont nettoyées des adaptateurs et les séquences de trop faible qualité, ou trop courtes après l'élimination des adaptateurs, sont éliminées.

Alignement. A priori, après avoir séquencé l'ADN, les nombreuses lectures obtenues peuvent provenir de n'importe quelle partie du génome, sauf en RRBS où les lectures ciblent certaines régions seulement. Afin de rechercher le niveau de méthylation de chaque cytosine détectée en WGBS et en RRBS, il faut aligner chaque lecture sur le génome de l'espèce, c'est-à-dire rechercher dans le génome la position d'origine d'une séquence similaire à celle de la lecture. Néanmoins, en raison du traitement au bisulfite et donc du changement de la base C si elle n'est pas méthylée, un outil d'alignement spécifique doit être utilisé. Parmi les outils existants d'alignement des données traitées au bisulfite, Bismark est l'outil le plus utilisé [155][165]. Il se base sur Bowtie qui est un aligneur de courtes séquences ultra-rapide et prenant peu de mémoire [166]. Il utilise une stratégie d'indexation basée sur la transformation de Burrows-Wheeler [167]. Chaque lecture est alignée sur le génome et le score d'alignement calculé est basé sur la similarité des séquences. Ce score augmente avec le nombre de bases en commun avec le génome de référence et diminue avec le nombre de substitutions, ou *mismatches*, et le nombre d'insertions/délétions, ou *gaps*. La qualité d'alignement, ou MAPQ (*Mapping Quality*), indique la fiabilité de la position sur laquelle se positionne la lecture. Plus l'écart entre le meilleur score d'alignement et le second meilleur score d'alignement est grand, plus le meilleur alignement peut être considéré comme unique, et plus haute devrait être la valeur du MAPQ [168]. Si le score d'alignement est élevé et le MAPQ faible, alors la lecture s'aligne parfaitement sur plusieurs positions. Au contraire, si le score d'alignement est faible et le MAPQ élevé, la lecture s'aligne avec des *mismatches* mais sa position dans le génome est beaucoup plus probable que n'importe quelle autre. À partir des données de séquence, Bismark réalise deux étapes (Figure 1.7). Tout d'abord, il convertit le génome de référence deux fois. Toutes les bases C sont converties en T et toutes les bases G sont converties en A. Puis il fait de même avec les lectures traitées au bisulfite. Ainsi, il exécute 4 processus d'alignement pour trouver le meilleur alignement possible, dont le résultat est présenté

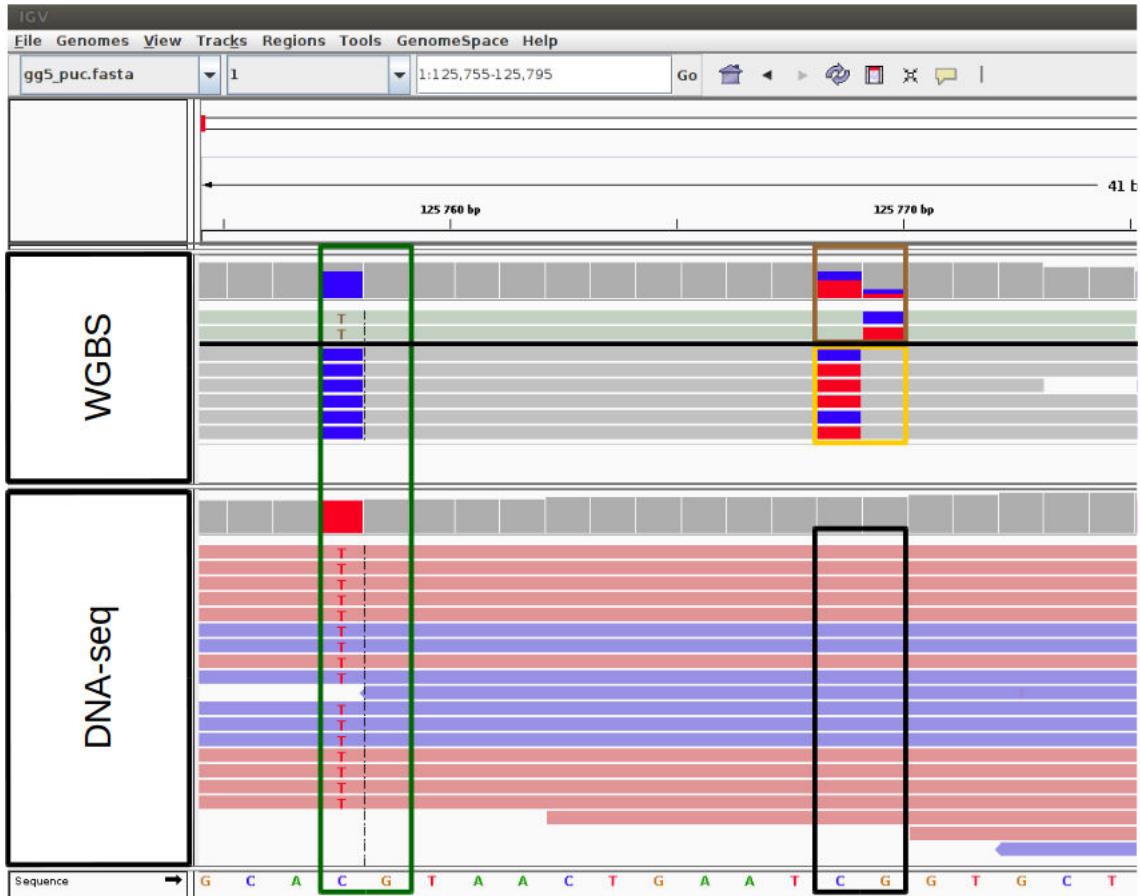


FIGURE 1.8 – Visualisation des lectures alignées sur la référence. Grâce au navigateur de génome IGV (*Integrative Genomics Viewer*), nous pouvons visualiser les lectures d'un échantillon traité au bisulfite (WGBS) et les lectures de ce même échantillon issues du DNA-seq, qui sont alignées sur le génome de référence de la poule (GCA_000002315.3). Sur les lectures WGBS, une ligne noire sépare les lectures provenant du brin sens (ou *forward*, brins OT et CTOT) de l'ADN (encadré orange) de celles provenant du brin anti-sens (ou *reverse*, brins OB et CTOB) de l'ADN (encadré marron) (cf Figure 1.6). Sur le CG, le taux de méthylation du C est calculé sur les lectures du brin sens, et l'encadré orange montre 6 positions dont 4C (carrés rouges) et 2T (carrés bleus), soit un taux de méthylation de 66,67%. De même, le taux de méthylation du G, par symétrie, est calculé sur les lectures en *reverse*, et l'encadré marron révèle 2 lectures : 1C + 1T, soit 50% de méthylation. Ainsi, le taux de méthylation du CG est de $(1+4)/(6+2) = 62,50\%$. Dans un second temps, on observe que la position de l'encadré vert est également un CG dans le génome de référence. Les lectures provenant du brin anti-sens (au dessus de la ligne noire) sont toutes potentiellement converties de G en A (cf Figure 1.6). Il ne s'agit pas d'une cytosine totalement déméthylée mais d'un SNP C/T, l'individu étant homozygote pour toutes les lectures. Cela est confirmé par la présence du SNP sur les lectures du DNA-seq (encadré vert). La cytosine voisine (encadré noir) n'a pas SNP sur les lectures non traitées au bisulfite.

avec la séquence d'origine (lecture issue du séquenceur et génome non converti). Ce meilleur alignement est décrit dans un fichier de format standard appelé SAM (*Sequence Alignment/Map*). Ce type de format est constitué d'une lecture par ligne et contient toutes les informations issues du séquençage et de l'alignement, notamment l'identifiant de la lecture, l'information du *paired-end* ou *single-end*, sa position dans le génome et le MAPQ de l'alignement. Ce type de fichier peut être analysé avec le logiciel SAMtools [169].

Suppression des biais d'alignement. La préparation de la librairie pour le séquençage haut-débit implique souvent une étape d'amplification pour augmenter la taille de la librairie, donc le nombre de chacun des fragments que l'on veut séquencer, afin de mieux détecter les séquences d'ADN. Cependant, l'amplification par PCR peut souvent introduire un biais dans la composition de l'échantillon, conduisant à une sous-représentation potentielle de certains fragments. Beaucoup de lectures identiques (même séquence) sont générées et vont s'aligner au même endroit sur le génome. Il est préférable de les enlever car elles augmentent artificiellement la profondeur et peuvent biaiser la mesure de la méthylation si toutes les amplifications n'ont pas la même efficacité. L'outil utilisé, SAMtools *rmDup*, par exemple, recherche les positions identiques des paires de lectures issues de l'alignement et ne conserve que la paire de lecture qui a la qualité d'alignement la plus élevée. Cependant, cette étape ne se fait pas en RRBS car pour une position donnée, les fragments digérés par MspI commencent par définition au même endroit.

Détection de la méthylation. Une fois les lectures alignées sur le génome de référence, pour chaque cytosine, il s'agit d'analyser la composition en bases dans les lectures à la position génomique correspondante. La détection du nombre de C et de T constitue la première étape des analyses de méthylation. De nombreux outils existent [162], et notamment le package R MethylKit qui permet d'analyser les profils de méthylation d'un génome entier [170]. Le taux de méthylation est calculé grâce à un nombre de C divisé par la somme de C+T en fonction des lectures présentes pour chaque position (Figure 1.8). Le pourcentage de lectures contenant un C à cette position donne le pourcentage de méthylation. À partir du fichier d'alignement, MethylKit détecte tous les C avec au moins une lecture. Les informations présentes pour chaque cytosine dans le fichier de résultats bruts sont le nom du chromosome, sa position, le brin sur lequel elle se trouve, la couverture et la fréquence de C et de T. La méthylation étant symétrique entre les brins d'ADN [171], l'information de méthylation est présente sur les deux brins : les niveaux de méthylation mesurés sur le C et le G voisin sont ajoutés pour obtenir un taux de méthylation par CG. Néanmoins, un problème dû à la présence de SNP peut apparaître (Figure 1.8). Quand l'individu analysé est hétérozygote C/T, voire homozygote T/T à une position C du génome de référence, le ratio $C/(C+T)$ ne représente pas un taux de méthylation, mais, en partie au moins, la présence de l'allèle T dans le génome, avant traitement au bisulfite.

Contrôle de la conversion au bisulfite. Pour certifier que toutes les cytosines non méthylées ont bien été transformées en thymines lors du traitement au bisulfite, il faut calculer un taux de conversion des cytosines hors des dinucléotides CG. En effet, si la conversion expérimentale n'a pas bien fonctionné, il reste des C non transformés en T alors qu'ils ne sont pas méthylés, ce qui entraîne une sur-estimation du taux de méthylation. Chez les vertébrés, dans la plupart des tissus, les C hors CG sont non méthylés donc doivent être transformés en T. Il est admis qu'un taux d'erreur acceptable doit être inférieur à 1% [172]. Il est également possible d'ajouter un contrôle interne, ou *spike-in*, pour savoir si la conversion au bisulfite a bien fonctionné.

Différentiel de méthylation. Afin de détecter une différence du taux de méthylation entre deux groupes de plusieurs individus, chaque groupe représentant une condition du dispositif expérimental, plusieurs stratégies existent, dont l'utilisation du package R DSS (*Dispersion Shrinkage of Sequencing data*) [173]. Il est basé sur un modèle bêta-binomial : la distribution binomiale capture la variation due à l'échantillonnage aléatoire des fragments d'ADN pendant le séquençage (la lecture peut être soit méthylée soit non méthylée, deux états possibles), et la distribution bêta modélise la variation du taux de méthylation (entre 0 et 100% de méthylation). Et enfin, une fonction de lissage, ou *smoothing*, utilise la corrélation spatiale entre les niveaux de méthylation des sites CpG voisins le long du génome [104][174], c'est-à-dire que les sites CpG proches sont assimilés à des pseudo-réplicats pour estimer une moyenne de méthylation du CpG en fonction de ses sites voisins et donc estimer une dispersion des valeurs de méthylation. Lorsque les sites CpG voisins détectés sont denses, comme en WGBS, le lissage permet d'obtenir de meilleures estimations des moyennes de méthylation. Néanmoins, en RRBS, l'option de lissage n'est pas recommandée en raison de la présence des sites CpG détectés à partir d'une fraction du génome seulement.

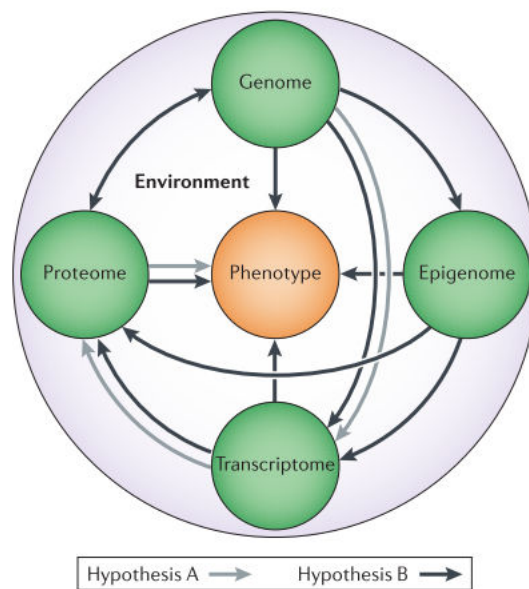


FIGURE 1.9 – Analyses intégratives de la variabilité phénotypique selon plusieurs types de données. Les flèches grises montrent l’hypothèse d’une approche multi-étapes, c’est-à-dire une méthode d’analyses hiérarchiques ou par étapes qui réduit l’information recherchée au fur et à mesure de l’intégration des données. Les flèches noires expriment l’idée d’une combinaison à tous les niveaux omiques, c’est-à-dire une intégration simultanée des informations des données pour produire des modèles avec plusieurs variables sur différentes échelles (figure issue de [177]).

1.3.3.2 Intégration de données -omiques

L'essor des nouvelles technologies étend le champs des données disponibles, telles que les données de séquençage du génome (DNA-seq, *DNA sequencing*), de l'épigénome (BS-seq, ChIP-seq ou *Chromatin Immunoprecipitation sequencing*), du transcriptome (RNA-seq, *RNA sequencing*) ou du protéome (spectrométrie de masse). L'association entre la variation phénotypique et la variation observée par l'analyse de ces données (expression des gènes, variation épigénétique, variation protéique) est de plus en plus explorée afin de mieux comprendre la complexité des processus biologiques. Combiner plusieurs types de données peut compenser les informations manquantes et/ou peu fiables d'un seul type de données. En outre, si la même information est recoupée dans plusieurs types de données, cela réduit le nombre de faux positifs. Les méthodes émergentes d'intégration de données ont pour but de combler le fossé entre la capacité de générer de grandes quantités de données, avec l'avènement des NGS, et la compréhension des systèmes biologiques. Il existe deux approches d'intégration de données : soit l'analyse multi-étapes, qui implique l'intégration de la donnée en utilisant une analyse par étape ou hiérarchique, soit l'analyse méta-dimensionnelle, qui intègre les types de données, en tant que variables, pour construire un modèle multi-varié associé à un phénotype donné [175][176].

Plusieurs hypothèses sur la variabilité moléculaire pouvant expliquer le caractère phénotypique qui en résulte ont été proposées par Ritchie et al (Figure 1.9) [177]. La première hypothèse est que la variation de l'ADN influe sur l'expression des gènes, eux-mêmes impactent les protéines pour finalement changer le phénotype. Pour cette hypothèse, l'approche multi-étapes est la plus puissante et la plus facilement interprétable pour l'intégration des données. Par exemple, il faut d'abord associer des SNP à un caractère phénotypique particulier, puis limiter la quantité de données épigénomiques à ces SNP particuliers, avant d'analyser ensemble ces deux types de données. L'autre hypothèse est que toutes les variations moléculaires interagissent entre elles de façon complexe et donc non linéaire. Privilégier cette hypothèse revient à combiner toutes les échelles des données, c'est-à-dire utiliser l'approche méta-dimensionnelle. Un exemple est celui de la combinaison des informations des CNV (*Copy Number Variation*, nombre de copies d'un fragment génomique), de la méthylation et des ARN non codants qui sont ensuite réduites au phénotype étudié. Ces différentes analyses d'intégration de données sont maintenant possibles grâce aux développements d'outils dédiés [177], bien que beaucoup reste à faire dans ce domaine.

1.4 Objectifs de la thèse

Anticiper l'impact de fluctuations environnementales de nature climatique ou alimentaire est un enjeu crucial dans les systèmes de productions animales, et plus particulièrement sur la volaille, source majeure de protéines pour l'alimentation humaine. Cette influence de l'environnement sur les phénotypes passe en partie par des phénomènes épigénétiques, mécanismes de régulation de l'expression des gènes, transmissibles par mitose et/ou méiose, sans modification de la séquence d'ADN [25]. Parmi ces phénomènes on trouve la méthylation de l'ADN, qui correspond à l'addition, réversible, d'un groupement méthyle sur une base de l'ADN. L'amélioration des techniques de séquençage haut-débit et l'utilisation d'outils bioinformatiques permet de mieux comprendre les interactions entre les épigénomes et l'expression des gènes dans les modèles animaux et chez les plantes. Cependant, les mécanismes de régulation de l'expression du génome par des phénomènes épigénétiques sont assez peu connus chez les espèces domestiques, en particulier chez les oiseaux. Cette connaissance est nécessaire pour, à terme, déterminer la contribution de la méthylation de l'ADN à la plasticité phénotypique et donc pour identifier des leviers potentiels d'amélioration de l'adaptabilité des animaux aux changements environnementaux. Pour répondre à cette problématique, mon sujet de thèse s'intéresse à l'analyse du méthylome chez la poule par séquençage haut-débit. Ce travail de thèse repose sur deux dispositifs expérimentaux réalisés chez la poule, avec des questions scientifiques complémentaires. Le premier projet a permis de caractériser la méthylation de l'ADN sur l'ensemble du génome d'embryons entiers et son lien avec l'expression des gènes, et de détecter la présence de la méthylation allèle-spécifique. À partir des données tout génome de ce projet, un pipeline d'analyses de méthylation optimisé et utilisable sur d'autres jeux de données a été développé. Le second projet, en cours, élargit ces analyses à l'étude du méthylome de la poule pondeuse en réponse à des variations de l'environnement : stress de chaleur et utilisation d'un aliment pauvre en énergie. Sur plusieurs tissus d'individus adultes, une sous-représentation du génome est séquencée et analysée avec le pipeline développé dans le premier projet.

1.4.1 Plan de la thèse

Ce travail de thèse s'appuie sur deux projets de recherche complémentaires. Le chapitre 1 "*Introduction*", développé précédemment, donne le contexte de recherche sur l'impact de l'environnement sur la variabilité des phénotypes caractérisé par les phénomènes épigénétiques, et notamment la méthylation de l'ADN. Celle-ci est de plus en plus étudiée avec précision grâce aux techniques de séquençage haut-débit et au développement d'outils bioinformatiques. Une revue que nous avons rédigée en collaboration est intégrée à cette partie. Le chapitre 2 "*Dispositifs, données et développement informatique*" reprend les matériels et les méthodes mis en œuvre pour répondre aux objectifs de la thèse à travers les deux projets. Les données de méthylome issues du séquençage tout génome et d'une sous-représentation du génome chez la poule sont analysées grâce à la mise en place de traitements bioinformatiques optimisés. Le chapitre 3 "*Analyses tout génome des données de méthylome*" présente les résultats et la discussion associée pour le premier programme de recherche : caractérisation de la méthylation de l'ADN d'embryons de poulet entiers, de son lien avec l'expression des gènes dans le promoteur et le corps du gène, et de la méthylation allèle-spécifique. Les analyses de ce chapitre font l'objet d'une publication en cours d'écriture. Le chapitre 4 "*Analyses du méthylome en réponse à un stress environnemental*" décrit les résultats, acquis à l'heure actuelle dans le second programme, qui analysent la réponse du méthylome à un stress à la chaleur et à l'utilisation d'un aliment moins énergétique chez la poule pondeuse grâce à la réutilisation du pipeline développé précédemment. La discussion de ces résultats est réalisée au cours de ce chapitre. Enfin, le chapitre 5 "*Synthèse, conclusion générale, perspectives*" réalise une synthèse des chapitres 3 et 4 qui permet de comparer les données de séquençage, puis de comparer un pipeline développé par ailleurs pour valider nos choix d'analyses. À partir de cette synthèse, une discussion plus générale permet de conclure quant aux objectifs de la thèse, et d'ouvrir sur les perspectives de ce travail de thèse.

Chapitre 2

Dispositifs, données et développement bioinformatique

2.1 Méthylome d’embryons de poule

Dans un contexte global d’analyses épigénétiques, le projet ANR (*Agence Nationale de la Recherche*) Epi-Bird s’intéressait notamment à l’existence de l’empreinte génomique parentale chez les Oiseaux. Un dispositif animal dédié a été mis en place chez la poule, permettant grâce à des croisements réciproques de discerner l’expression allèle-spécifique de l’empreinte génomique parentale (cf partie 1.2.4 page 39).

Les données utilisées sont issues du dispositif animal décrit dans la publication (cf partie 3.1 page 83). Brièvement, 8 embryons (à 4,5 jours de leur développement) sont obtenus à partir d’un croisement réciproque de deux lignées (Tableau 2.1). Ces deux lignées de poule (R^- et 6) sont les plus consanguines et les plus distantes génétiquement possible. Elles ont été choisies pour pouvoir détecter un grand nombre de polymorphismes, et déduire avec la plus grande certitude possible l’origine parentale de chaque allèle pour les positions hétérozygotes chez les descendants. À partir de la lignée R^- , décrite ci-après (cf partie 2.2 page 75) et de la lignée 6, sélectionnée pour sa résistance aux maladies comme la Salmonellose [178], on dispose de deux mâles et deux femelles, chacun utilisé pour un sens de croisement : $\text{♀ } 6 \times \text{♂ } R^-$ et $\text{♀ } R^- \times \text{♂ } 6$ (Figure 2.1). Des embryons très hétérozygotes sont obtenus sur lesquels l’origine parentale du polymorphisme peut être connue. Grâce à cette forte hétérozygotie, un niveau d’expression associé à chaque allèle (ASE) peut être analysé sur un grand nombre de gènes. Le dispositif a ainsi permis de démontrer l’absence d’empreinte génomique chez la poule, au moins dans les embryons totaux de 4,5 jours [131]. Lors de l’étude de la méthylation allèle-spécifique décrite ci-après (cf partie 3.2.4 page 119), la lignée 6 est représentée par les parents 7127 et 7128, tandis que la lignée R^- est représentée par les parents 8303 et 8403. Les croisements dont sont issus les embryons proviennent soit de $7127 \text{ ♀} * 8403 \text{ ♂}$ ($= \text{♀ } 6 \times \text{♂ } R^-$), soit de $8303 \text{ ♀} * 7128 \text{ ♂}$ ($= \text{♀ } R^- \times \text{♂ } 6$). Chaque croisement est noté par la suite avec le nom de la mère : croisement 7127 pour $7127 \text{ ♀} * 8403 \text{ ♂}$, et croisement 8303 pour $8303 \text{ ♀} * 7128 \text{ ♂}$. Les noms des embryons sont notés par le nom de leur mère (exemple : 7127emb08 est issu d’une mère 7127 croisée avec le père 8403).

Pour chaque embryon, trois types de données sont disponibles :

- Données de transcriptome : produites par RNA-seq et analysées par ailleurs [131], elles ont permis d’attribuer à chaque gène un niveau d’expression dans chacun des embryons.
- Données de génome : produites par DNA-seq et analysées par ailleurs [179], elles ont permis de révéler les variations propres à chaque individu (SNP).
- Données de méthylome : produites par WGBS, elles permettent d’analyser la méthylation de l’ADN sur tout le génome, à l’échelle de la base. Elles sont au cœur de ce travail de thèse, en lien avec les deux jeux de données précédents. L’objectif est de caractériser le méthylome du poulet sur un dispositif adapté pour étudier, notamment, les éventuels phénomènes allèle-spécifiques.

Les données WGBS sont décrites dans la publication (cf partie 3.1 page 83). Un séquenceur Illumina

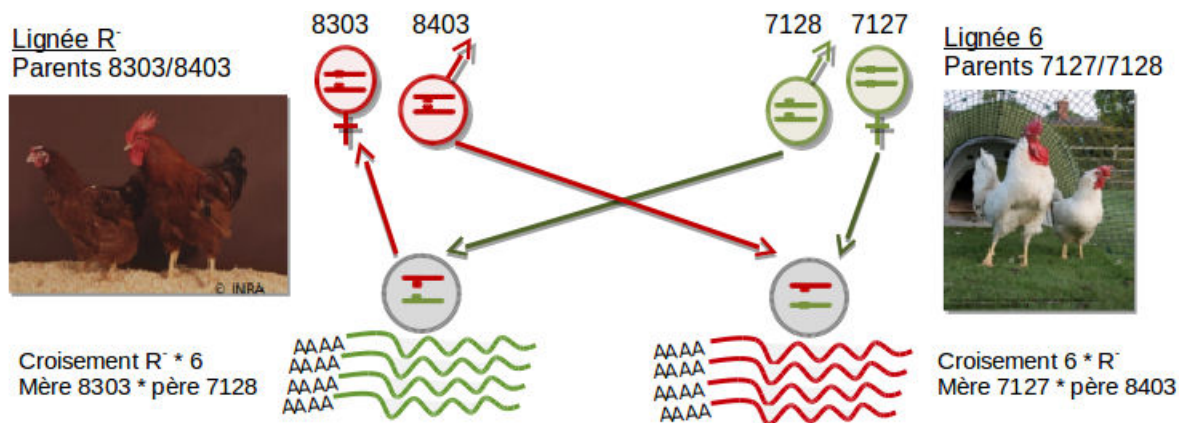


FIGURE 2.1 – Schéma du dispositif expérimental du projet EpiBird. Le dispositif est constitué de deux croisements réciproques de deux lignées de poules : 6 et R⁻. Quatre embryons sont issus d'un croisement ♀ 6 (mère 7127) x ♂ R⁻ (père 8403) et quatre embryons proviennent d'un croisement ♀ R⁻ (mère 8303) x ♂ 6 (père 7128).

Nom embryon	Sexe embryon	Mère	Lignée mère	Père	Lignée père
7127emb08	femelle	7127	6	8403	R ⁻
7127emb15	mâle	7127	6	8403	R ⁻
7127emb20	mâle	7127	6	8403	R ⁻
7127emb23	femelle	7127	6	8403	R ⁻
8303emb12	femelle	8303	R ⁻	7128	6
8303emb14	mâle	8303	R ⁻	7128	6
8303emb18	mâle	8303	R ⁻	7128	6
8303emb22	femelle	8303	R ⁻	7128	6

Tableau 2.1 – Dispositif du projet EpiBird (données WGBS). Pour chacun des 8 embryons, leur sexe et leur croisement sont décrits, ainsi que le nom de leurs parents.

Lignée	Tissu	Contrôle	Avant stress	Stress alimentaire	Stress thermique	Nombre total
R ⁻	Hypothalamus	6	0	6	6	18
R ⁺	Hypothalamus	6	0	6	6	18
R ⁻	Foie	6	0	6	6	18
R ⁺	Foie	6	0	6	6	18
R ⁻	Sang	6	6	6	6	24
R ⁺	Sang	6	6	6	6	24
Total		36	12	36	36	120

Tableau 2.2 – Dispositif du projet ChickStress (données RRBS). Ce tableau récapitule le nombre d'échantillons par lignée et pour chaque tissu dans chaque condition expérimentale, pour les 36 individus étudiés.

HiSeq2000 a été utilisé pour générer des lectures *paired-end* 2 x 101 pb. Environ 161 millions de fragments d'ADN par embryon ont été obtenus, soit 81 Go de données.

2.2 Réponse du méthylome à un stress environnemental

L'impact de l'environnement sur la variabilité des phénotypes est modulé par des phénomènes épigénétiques qui modifient l'expression des gènes. Ces marques épigénétiques répondent aux stress abiotiques mais les mécanismes moléculaires sous-jacents restent peu connus. Parmi les stress dus à l'environnement, les effets de la température et de régimes alimentaires carencés sont très étudiés en raison de leur importance dans les productions animales (cf partie 1.1.1 page 23). Le projet ANR ChickStress, dans lequel s'inscrit ma thèse, rassemble un large éventail de caractères connus comme étant affectés par ces stress (vitesse de ponte, poids des œufs, composition corporelle, ...) ou importants dans la tolérance à la chaleur et l'efficacité alimentaire (ingestion alimentaire, couverture des plumes, thermogenèse induite par le régime,...). Chez la poule, toutes les lignées ne répondent pas de la même façon aux stress environnementaux. Certaines lignées sont plus tolérantes à la chaleur mais souvent moins productives en conditions standards, tandis que d'autres produisent plus dans un environnement tempéré [6]. De même, optimiser le régime alimentaire des poules est un enjeu pour les filières avicoles, notamment dans le but d'améliorer les caractères de production dans un contexte mondial tendu en matières d'alimentation animale. Pour comprendre les mécanismes moléculaires à l'origine de la variabilité des caractères d'adaptation des différentes lignées, il est nécessaire d'utiliser des lignées de poules qui divergent pour des critères étudiés. L'étude de l'adaptabilité de ces lignées à leur environnement peut passer par la caractérisation du méthylome. Chaque méthylome dépend du type cellulaire étudié. Le foie et l'hypothalamus sont les organes majeurs pour l'homéostasie énergétique, c'est-à-dire qu'ils sont nécessaires pour maintenir un équilibre entre l'énergie issue de la prise alimentaire et l'énergie dépensée pour le maintien de la stabilité de l'organisme [180]. L'hypothalamus induit une réponse en cas de stress par des cascades hormonales pour maintenir l'homéostasie [181] (cf partie 1.1.1 page 25). Il est le centre de la réponse au stress et contrôle la faim, la soif comme la température corporelle. Le foie est également impliqué en cas de réponse au stress car il s'agit d'un carrefour métabolique, en particulier pour l'énergie à travers le métabolisme des lipides [182] (cf partie 1.1.2 page 27). De plus, comme vu précédemment (cf partie 1.1.1 page 25), des paramètres sanguins permettent de déceler les réponses à un stress à la chaleur. Comparer les réponses de l'épigénome aux stress environnementaux intra- et inter-tissulaires peut permettre de révéler les mécanismes moléculaires en jeu dans l'adaptation. Ainsi, le projet ChickStress propose d'étudier dans quelles mesures un stress thermique et un stress alimentaire chez l'oiseau adulte impactent le phénotype, et quels sont les mécanismes moléculaires impliqués. Notre implication dans le programme concerne les études de méthylome.

Dans le cadre de la thèse, le dispositif utilisé est le suivant. Chez la poule pondeuse, deux lignées expérimentales ont été sélectionnées de façon divergente sur le critère de la consommation alimentaire résiduelle [183] :

- R⁺ avec une forte consommation résiduelle : fort appétit, thermogenèse alimentaire importante, faible quantité de gras dans le corps, plus tolérant à la chaleur
- R⁻ avec une faible consommation résiduelle : faible appétit, faible thermogenèse, plus grande quantité de gras dans le corps, moins tolérant à la chaleur. Cette lignée est également utilisée dans le premier projet.

Des prélèvements de tissus ont été réalisés à l'abattage (33 semaines d'âge) sur 12 animaux contrôles (6 par lignée), 12 animaux stressés à la chaleur et 12 animaux stressés à un régime sub-optimal (6 par lignée par stress) (Tableau 2.2). Le premier stress consiste, à partir de la dix-septième semaine d'âge, à nourrir les individus avec des aliments moins énergétiques comme le colza et le blé au lieu du maïs et du soja jusqu'à l'abattage à 33 semaines. Parallèlement, un autre groupe est soumis à un stress de chaleur qui consiste à augmenter de 10°C (de 22°C à 32°C) l'exposition des poulets entre la vingt-huitième et la trente-troisième semaine d'âge. Les prélèvements du sang total, de l'hypothalamus et du foie se font à l'abattage pour les groupes stressés et contrôle. Du sang est également prélevé avant le stress pour le groupe stressé à la chaleur. Le sang total permet également d'effectuer des numérations sanguines.

2.2.1 Protocole expérimental RRBS

Le protocole qui a été utilisé pour produire les données RRBS est décrit dans la figure 2.2.

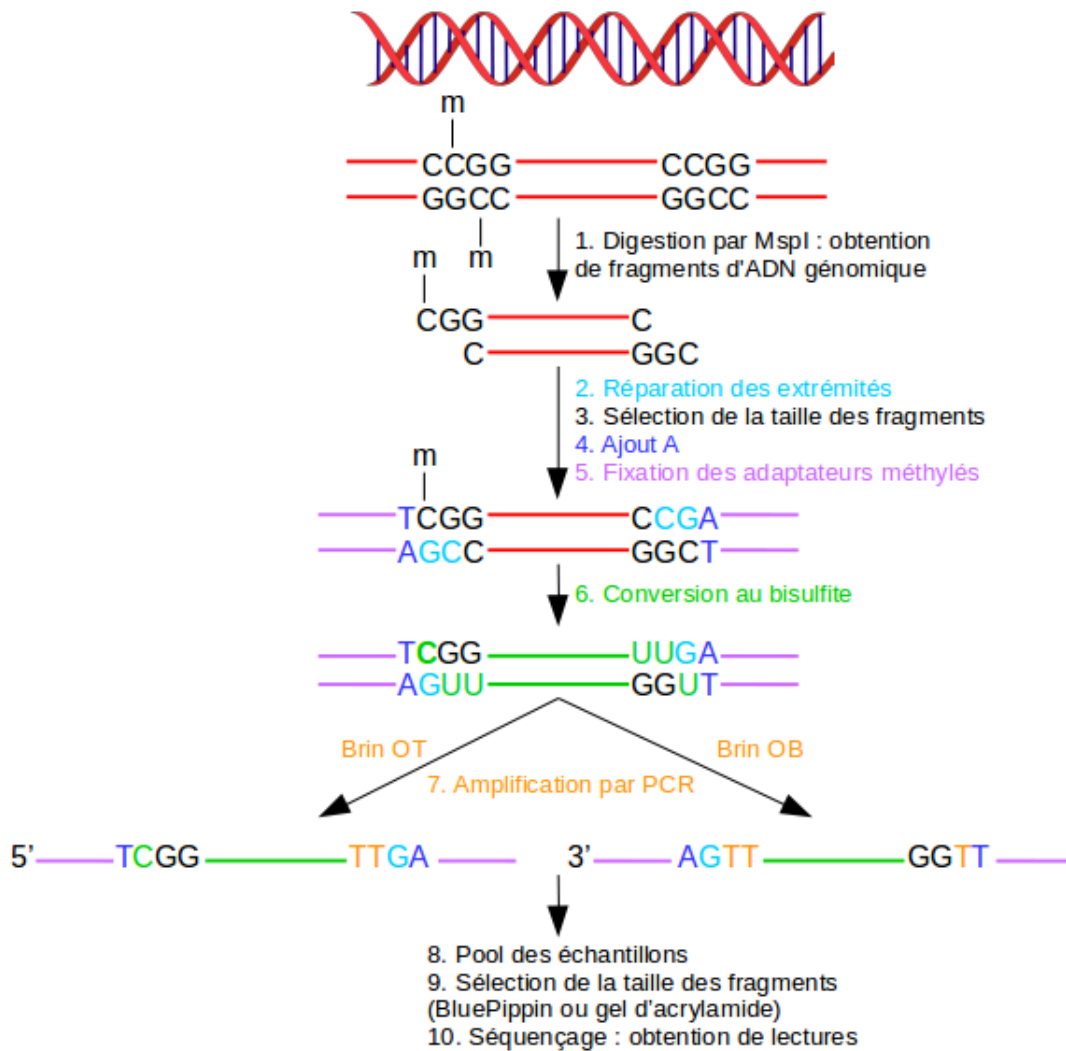


FIGURE 2.2 – Étapes du séquençage RRBS (Protocole utilisé : NEXTFLEXTM Bisulfite-Seq Kit, www.catalogue.proteigene.com). L'ADN génomique est digéré par *MspI* qui génère des extrémités cohésives. Ces extrémités sont "réparées" (extrémités à bouts collants) : ajout des nucléotides complémentaires. Seuls les petits fragments sont récupérés (de 100 à 600 pb). Puis, une adénosine est ajoutée sur les deux brins de l'ADN en 3'. Cette base sert à lier les adaptateurs qui sont ensuite ajoutés. Ce sont de petites séquences qui servent à identifier les fragments d'ADN génomique et à réaliser une matrice pour l'amorçage de l'amplification. Les adaptateurs totalement méthylés (uniquement des 5mC) empêchent la désamination des C lors de la conversion au bisulfite, réalisée avant l'amplification. Ces fragments sont ensuite amplifiés par PCR. Tous les individus sont marqués de façon spécifique puis regroupés pour avoir la même quantité d'ADN par échantillon dans le même tube (multiplexage). Une sélection des fragments de taille souhaitée est réalisée soit grâce au BluePippin (www.sagescience.com/products/bluepippin/) soit sur gel d'acrylamide. Avec la technique du BluePippin, l'ADN est déposé dans le puits puis par migration, les fragments plus petits ou plus grands que la taille sélectionnée sont éliminés. Seuls les fragments d'ADN d'intérêt sont récupérés. Cette sélection de taille peut également se faire sur gel d'agarose ou d'acrylamide. Après migration sur le gel, il est découpé au niveau des fragments de taille souhaitée (gel d'acrylamide plus fin que le gel d'agarose donc les fragments sont mieux séparés). Ensuite, l'ADN est récupéré après une purification sur billes puis séquencé.

Le séquenceur utilisé peut lire un minimum de 250 millions de fragments par ligne. La technique de RRBS génère des fragments d'ADN qui commencent tous par la même suite de bases, correspondant au site de restriction dont ils sont tous issus. Pour analyser correctement ces données particulières, le séquençage nécessite l'ajout dans la ligne de séquence d'une quantité importante (15% à 20%) de fragments provenant d'une banque classique, dont les débuts de séquence sont très divers. Seuls environ 200 millions de lectures correspondront donc aux fragments RRBS, soit, pour 12 individus, autour de 17 millions de fragments par individu. Pour une analyse très fiable du taux de méthylation, nous visons une profondeur de 50X, issue de la répartition sur le génome des 17 millions de séquences qui seront disponibles. Il nous faut donc un maximum d'environ 340 000 fragments génomiques différents à séquencer par individu (17M / 50X). Le taux d'alignement moyen sur le génome de référence en RRBS est en moyenne d'environ 60%, parfois plus pour les génomes les mieux assemblés (63-75% sur les données humaines [184]; 60-64% chez le porc et la chèvre [185][186]). Il faut donc multiplier par 60% le nombre de fragments à analyser pour s'assurer d'une bonne profondeur de séquence sur les régions analysées. Nous souhaitons donc sélectionner environ 204 000 fragments par individu.

La sélection de la taille des fragments doit permettre de réunir un maximum de CG lors de la détection avec le séquençage d'un minimum de fragments, pour optimiser les coûts. Pour connaître la taille voulue, une digestion enzymatique *in silico* du génome de référence est effectuée. Plusieurs enzymes peuvent être choisies. L'enzyme MspI a un site de restriction CCGG et digère indifféremment les sites méthylés ou non. L'utilisation de cette enzyme permet d'obtenir des fragments riches CG de différentes tailles. Un script d'analyses permet de compter le nombre de fragments digérés, le pourcentage du génome couvert et le nombre de CG par fragment est calculé. Ces fragments sont ensuite regroupés par intervalle de taille. Cette digestion *in silico* conditionne la suite des analyses. Les CG détectés sur les petits fragments sont les plus nombreux (Figure 2.3). Le séquençage étant effectué en lectures de 50 pb, les fragments beaucoup plus courts perdent leur intérêt. On choisit donc de fixer la limite inférieure de sélection des fragments à une taille de 40 pb [155]. Différentes fenêtres de taille de fragments sont obtenues avec un nombre de CG théoriquement détectés dans le génome (Tableau 2.3). Ainsi grâce aux intervalles de fragments digérés de la référence et à la connaissance du nombre de fragments obtenus par individu, on déduit que l'intervalle qui se rapproche le plus de 204 000 fragments est celui de 40-100 pb (Figure 2.4). Donc la sélection des fragments compris entre 40 et 100 pb est la plus optimale pour un séquençage 1 x 50 pb, avec un nombre de lignes de séquenceur ne permettant pas un multiplexage moins dense que 12 échantillons par ligne. Cette approche reste intéressante en terme d'efficacité car on doit obtenir des données analysables pour 6,4% des CG du génome en n'ayant séquencé que 1% du génome. Le séquenceur Illumina HiSeq 4000 est utilisé pour obtenir des lectures *single-end* 1 x 50 pb (plateforme GenomEast, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg). Les données correspondantes à ce projet sont encore en cours d'acquisition. Seules les données du tissu hypothalamus ont été générées pendant ma thèse, ce qui représente 21,49 millions de fragments par individu, soit 16 Go de données (Tableau 2.5).

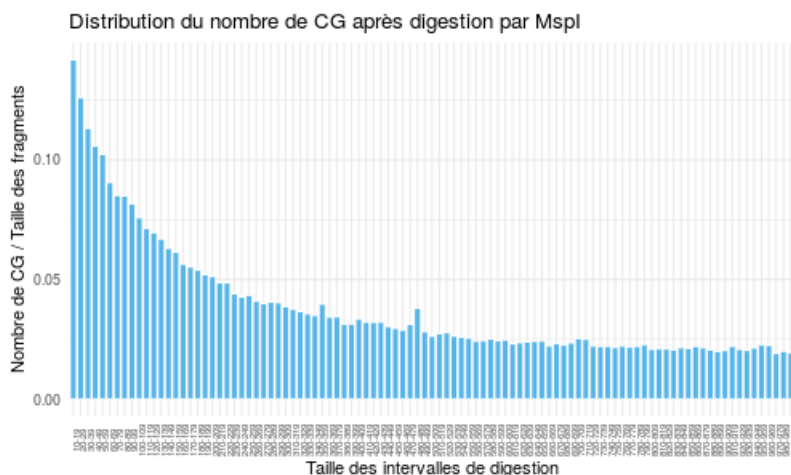


FIGURE 2.3 – Distribution des CG dans les fragments digérés par MspI. Le nombre de CG est divisé par la taille des fragments du génome *Gallus gallus* digéré par MspI, qui sont regroupés par fenêtre de 10 pb, de 10 à 1000 pb, avec un comptage de fragments tous les 10 pb.

Fenêtres	Nb fragments	Taille fragments	% Génome couvert	Nb CG	% CG couverts
40-50 pb	43742	1942468	0,16	205512	1,15
40-90 pb	170105	10564733	0,86	982402	5,49
40-100 pb	191092	12545363	1,02	1143720	6,39
40-120 pb	226046	16363765	1,33	1424163	7,96
40-140 pb	254544	20042001	1,63	1674506	9,36

Tableau 2.3 – Fenêtres de taille des fragments. Pour chaque fenêtre testée, le nombre de fragments, leur taille, le pourcentage de génome couvert, le nombre de CG détectés et leur couverture sur le génome sont récapitulés dans ce tableau.

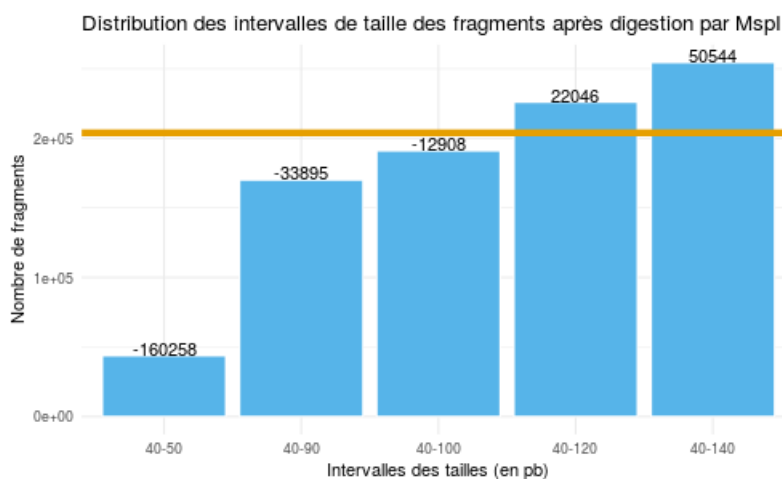
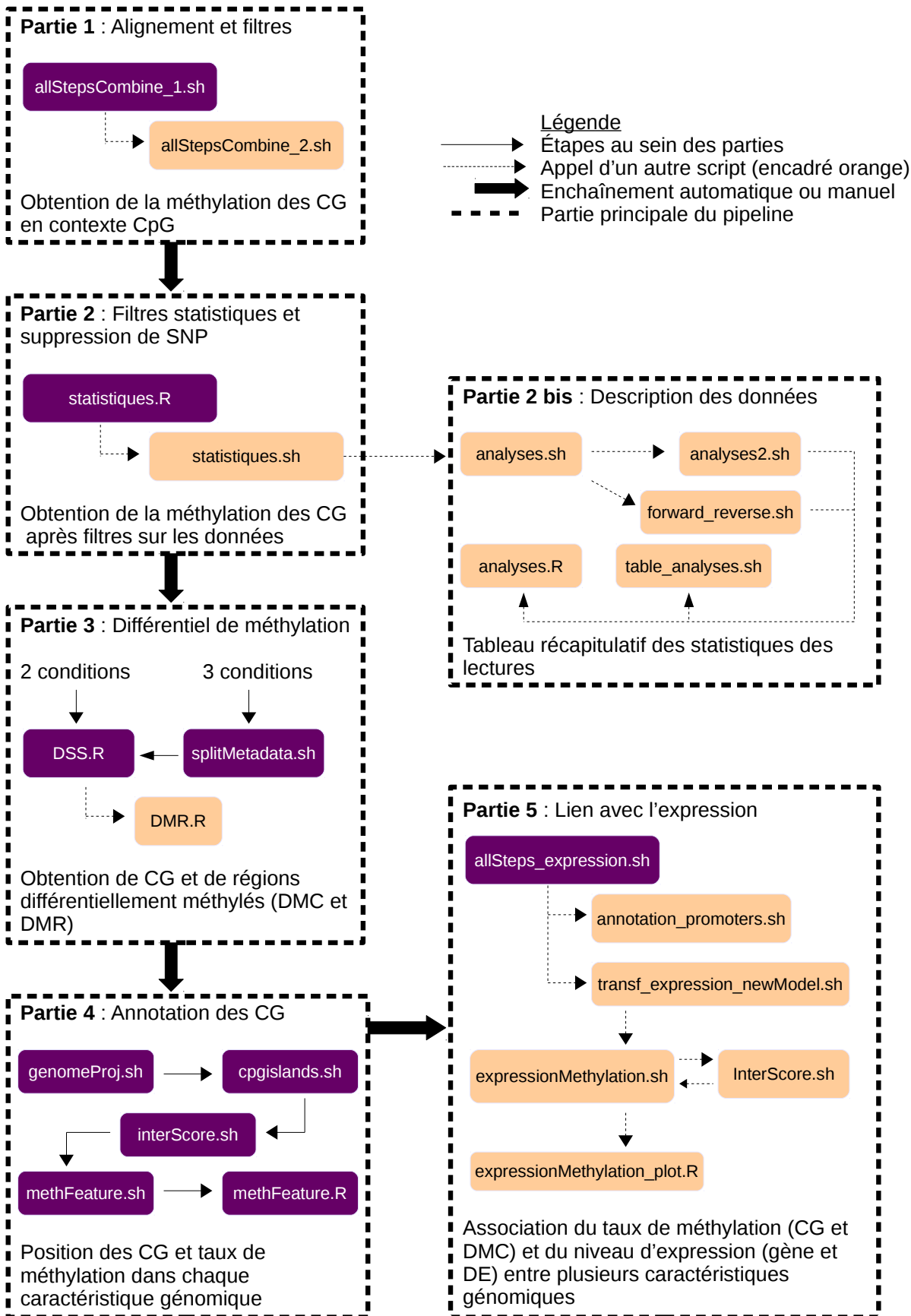


FIGURE 2.4 – Distribution des intervalles de taille des fragments après digestion enzymatique. Les fragments digérés par MspI sont regroupés par intervalles de 10 à 100 pb. La ligne orange représente les 204 000 fragments issus de la digestion *in silico* que l'on doit atteindre. Les chiffres au-dessus de chaque barre d'histogramme sont la différence entre le nombre de fragments détectés et les 204 000 fragments digérés *in silico*.

Échantillons	Lignée	Stress
H_004_0966_Rp_FS	R ⁺	Alimentaire
H_006_0978_Rp_FS	R ⁺	Alimentaire
H_007_0992_Rp_FS	R ⁺	Alimentaire
H_013_1140_Rp_FS	R ⁺	Alimentaire
H_014_1007_Rp_FS	R ⁺	Alimentaire
H_051_0777_Rm_CT	R ⁻	Contrôle
H_062_1006_Rp_CT	R ⁺	Contrôle
H_065_0629_Rm_CT	R ⁻	Contrôle
H_088_0935_Rm_CT	R ⁻	Contrôle
H_099_0602_Rm_CT	R ⁻	Contrôle
H_108_1122_Rp_CT	R ⁺	Contrôle
H_132_0972_Rp_CT	R ⁺	Contrôle
H_161_0835_Rm_CT	R ⁻	Contrôle
H_168_1060_Rp_CT	R ⁺	Contrôle
H_185_1150_Rp_CT	R ⁺	Contrôle
H_197_0815_Rm_FS	R ⁻	Alimentaire
H_202_0624_Rm_FS	R ⁻	Alimentaire
H_203_0754_Rm_FS	R ⁻	Alimentaire
H_204_0920_Rm_FS	R ⁻	Alimentaire
H_208_0827_Rm_FS	R ⁻	Alimentaire
H_210_0592_Rm_FS	R ⁻	Alimentaire
H_331_1020_Rp_HS	R ⁺	Chaleur
H_341_1094_Rp_HS	R ⁺	Chaleur
H_342_0819_Rm_HS	R ⁻	Chaleur
H_357_0598_Rm_HS	R ⁻	Chaleur
H_375_0933_Rm_HS	R ⁻	Chaleur
H_410_0984_Rp_HS	R ⁺	Chaleur
H_417_1121_Rp_HS	R ⁺	Chaleur
H_427_0739_Rm_HS	R ⁻	Chaleur
H_446_0610_Rm_HS	R ⁻	Chaleur
H_463_1077_Rp_HS	R ⁺	Chaleur
H_470_1055_Rp_HS	R ⁺	Chaleur

Tableau 2.5 – Échantillons d’hypothalamus d’individus adultes issus des données RRBS. Ce tableau représente les individus étudiés avec leur lignée (Rm = R⁻, Rp = R⁺) et le stress environnemental auquel l’individu a été soumis : HS = *Heat Stress* = stress chaleur ; FS = *Feed Stress* = stress alimentaire. Les individus contrôles sont notés CT.

FIGURE 2.5 – Schéma des enchaînements de scripts du pipeline.



2.3 Développement d'un pipeline bioinformatique

Afin d'analyser ces données de méthylome issues du séquençage haut-débit, un pipeline d'analyses et des scripts complémentaires ont été mis en place. Développé avec le langage bash pour une meilleure manipulation de fichiers volumineux, et en langage R pour l'optimisation des analyses statistiques et la visualisation sous forme de graphiques, le pipeline comprend plusieurs parties qui s'articulent ensemble, mais dont chacune peut être exécutée séparément (Schéma 2.5). Le choix de la ligne de commande a été fait pour la simplicité d'utilisation et la facilité de mise en oeuvre sur un cluster de calcul. Chaque logiciel utilisé est intégré dans l'enchaînement des scripts. Afin d'avoir les ressources informatiques nécessaires au développement des scripts, le cluster de calcul de la plateforme INRA a été utilisé (`bioinfo.genotoul.fr`). Un cluster de calcul est un ensemble d'ordinateurs dont les ressources informatiques sont regroupées. Cela forme une machine parallèle dédiée au calcul intensif. Durant une grande partie de ma thèse, le cluster de calcul dédié était utilisé sous SGE (*Sun Grid Engine*, genotoul), gestionnaire des ressources qui gère l'utilisation des noeuds de calcul. Une mise à jour récente dans le changement du gestionnaire de ressources (utilisation de SLURM, *Simple Linux Utility for Resource Management*, genologin) a nécessité que j'adapte mon pipeline d'analyses. L'analyse des données se fait sur un certain nombre d'individus et le pipeline a été construit de façon à limiter le plus possible les ressources informatiques et le temps de calcul. La parallélisation des individus permet de diviser le temps de calcul et de partager la mémoire utilisée.

La première partie, des données brutes à la détection des positions de CG "brutes", comprend le nettoyage des données avec le logiciel Trim Galore [155], l'alignement avec Bismark [155], la suppression des duplicats de PCR avec *samtools rmdup*, et enfin la détection de la méthylation de l'ADN des cytosines en contexte CpG avec le package MethylKit [170]. Ces étapes sont toutes parallélisées par échantillon. Pour ne pas confondre les éventuels biais expérimentaux et les effets biologiques réels de l'individu, les données peuvent être séquencées sur plusieurs lignes du séquenceur et donc plusieurs fichiers sont obtenus. Dans le premier projet, 6 embryons sont issus de 3 lignes de séquençage et 2 sur une seule ligne. En RRBS, l'hypothalamus a été séquencé sur deux ligne de séquençage pour chaque échantillon. Chacun de ces réplicats techniques est traité individuellement et donc en parallèle jusqu'à l'étape d'alignement. En effet, les étapes de nettoyage des données (suppression des adaptateurs) et l'alignement sont appliquées à chaque lecture, indépendamment des autres lectures. Néanmoins, la suppression des duplicats doit se faire avec toutes les données par individu en raison de l'amplification unique effectuée au moment de la préparation de la librairie. Le choix est donc fait de fusionner les réplicats techniques après l'alignement. Ce choix a également permis de gagner en rapidité en terme de temps de calcul (15h35 lorsque la fusion des réplicats est avant l'alignement et 13h40 lorsque c'est après alignement, pour les mêmes ressources informatiques utilisées). En outre, le logiciel Bismark intègre une option de parallélisation de l'alignement des lectures. Bowtie, l'aligneur sur lequel se base Bismark, utilise deux processus (ou *thread*), ce qui permet de paralléliser l'alignement et donc de l'effectuer plus rapidement. La dernière étape de cette partie n'intègre pas de possibilité de parallélisation au sein du package. Afin de conditionner le bon enchaînement de l'analyse pour le bon échantillon, le fichier d'entrée principal de cette partie est un fichier de métadonnées. Il regroupe les informations nécessaires du *design* expérimental, c'est-à-dire les chemins d'accès aux fichiers de séquençage, l'espèce étudiée, les identifiants des individus, la condition de l'expérience, le numéro de la ligne de séquençage et en *paired-end*, le sens des lectures (sens ou *read 1*, et anti-sens ou *read 2*).

La deuxième partie filtre les CG "bruts" par une normalisation par la taille de la librairie, un minimum et maximum de couverture, récupère les CG communs entre les échantillons et supprime les SNP si un fichier est donné dans les paramètres d'entrée. Le fichier texte final, facilement manipulable, est la liste des CG avec leur taux de méthylation qui sont étudiés dans les analyses suivantes. Les scripts des parties suivantes sont des enchaînements de commandes bash, et de logiciels spécialisés dans la manipulation des données génomiques comme bedtools [187]. Chaque partie du pipeline utilise des ressources informatiques et un temps de calcul différent (Tableau 2.6). De plus, ce pipeline a pu être utilisé sur d'autres données WGBS chez la caille et une autre version des parties 1 et 2 a été implémentée sous Jflow, gestionnaire de pipelines développé à l'INRA [188] pour inclure des modèles plantes.

Les différents paramètres d'entrée ont été adaptés et optimisés à chaque étape du pipeline, pour les données issues du RRBS ou du WGBS. Ces paramètres sont discutés dans la partie 5.1 page 153.

Parties	Temps CPU	Mémoire	Nb nœuds
1. Alignement et filtres	4j 12h 29min	8 Go	8
2. Filtres statistiques et suppression SNP	6h 1min	8 Go	8
2bis. Description des données	2h 23min	5 Go	8
3. Différentiel de méthylation	2h 10min	8 Go	8
4. Annotation des CG	2h 6min	8 Go	8
5. Lien avec l'expression	40min	8 Go	4

Tableau 2.6 – Ressources informatiques utilisées par le pipeline. Pour chaque partie du pipeline, il est noté les ressources informatiques utilisées, comme la mémoire et le nombre de nœuds, et le temps de calcul (temps CPU).

Chapitre 3

Analyses tout génome des données de méthylome

L'objectif de cette première analyse était de caractériser le méthylome et les mécanismes de régulation de l'expression des gènes associés, mécanismes encore peu connus chez la poule malgré quelques études récentes [189][190][191]. Pour cette étude, nous disposions d'un dispositif animal construit pour la détection de l'empreinte génomique parentale, idéal pour détecter la méthylation allèle-spécifique. Grâce aux données de génome et de transcriptome disponibles sur ce dispositif, en plus du méthylome, nous avons pu effectuer une analyse complète intégrant les différents types de données. Les résultats de ces travaux sont décrits dans la publication suivante, en cours de préparation pour une soumission dans le journal *Scientific Reports*. Les analyses qui suivent n'ont pas été intégrées dans l'article mais sont pleinement intégrées à ce travail de thèse. Elles sont donc décrites dans la partie 3.2 page 109.

3.1 Projet de publication

Methylome analyses in chicken embryo: evidence of allele-specific methylation with no parent-of-origin effect

Marjorie Mersch¹, Sophie Leroux¹, Céline Noirot², Diane Esquerré^{1,3}, Sarah Djebali¹, Mireille Morisson¹, Gérald Salin⁴, Vincent Coustham⁵, David Gourichon⁶, Laure Frésard⁷, Sandrine Lagarrigue⁸, Sylvain Foissac^{1,+,*}, and Frédérique Pitel^{1,+,*}

¹UMR INRA / INPT ENSAT / INPT ENVIT, 24, chemin de Borde Rouge, 31326 Castanet-Tolosan, France

²INRA, UR875 MIAT, 24, chemin de Borde Rouge, 31326 Castanet-Tolosan, France

³INRA, UR875 Biométrie et Intelligence Artificielle, Castanet-Tolosan, France

⁴INRA, GeT-PlaGe Genotoul, 24, chemin de Borde Rouge, 31326 Castanet-Tolosan, France

⁵INRA, UMR Biologie des Oiseaux et Aviculture, F-37380 Nouzilly, France

⁶PEAT, INRA Tours, 37380 Nouzilly, France

⁷Department of Pathology, Stanford, CA 94305, USA

⁸Agrocampus Ouest, INRA, UMR1348 PEGASE, Rennes, France

*corresponding: frederique.pitel@inra.fr

+these authors contributed equally to this work

ABSTRACT

Advances in high-throughput technologies have enabled the genome-wide analysis of DNA methylation in relation with gene expression in many species. Here we took advantage of an animal design built to answer the question of the existence of genomic imprinting in chicken to gain insight into the chicken embryo methylome and its relationship with gene expression. By using DNA-Seq, RNA-Seq and BS-Seq data, we characterized the chicken methylome in whole embryos. We confirmed the distinctive features of vertebrates methylomes as the low methylation level of CG islands and the negative correlation between promoter methylation and gene expression. We demonstrate the existence of allele-specific methylation, and its association with allele-specific expression at several loci. We confirm the lack of genomic imprinting in chicken embryos at the methylation level.

Introduction

Epigenetic mechanisms take part in the regulation of gene expression and may be particularly sensitive to the environment¹. These mechanisms do not alter the DNA sequence composition and may be transmitted through mitosis or meiosis². Epigenetic landscape is composed of several molecular processes, such as DNA methylation or histone modifications. DNA methylation can be altered in response to environmental factors, showing an interaction between the variability of gene expression and environmental changes². Establishment of epigenetic marks are essential during the development³, when their sensitivity to environmental variation reaches its highest level⁴. Moreover, abnormal changes in DNA methylation status during embryonic development can lead to alterations in the phenotypes^{5,6}.

However, despite the importance of the role of DNA methylation in the regulation of gene expression, very little information exist about the underlying molecular mechanisms in chicken.

Here we took advantage of an animal design built to answer the question of the existence of genomic imprinting in chicken⁷, consisting in embryos from two reciprocal crosses between two different chicken lines. The design was structured to maximise the number of heterozygous genomic positions for which the parental origin of the allele can be undoubtedly identified. We were thus able to analyse Allele-Specific Methylation (ASM). Allele-specific methylation is a genomic phenomenon where different levels of methylation are observed between alleles of a same locus. ASM occurs in mammals either in parent of origin-dependant phenomena, such as genomic imprinting or X inactivation, or in sequence-dependant mechanisms^{8,9}. Sequence-dependant ASM involves SNPs that affect the level of CpG methylation¹⁰. The analysis of ASM events near expressed genes show a

high correlation with Allele-Specific Expression¹¹. We analyzed DNaseSeq, RNASeq and WGBS (Whole Genome Bisulfite Sequencing) data to decipher the link between DNA methylation and gene expression, including differences observed between sexes and cross ways. We further analyzed allele-specific methylation and allele-specific expression, and their interactions.

Results and discussion

The dataset used in this study is publically available under the SRA study accession number SRP033603. The mean number of reads obtained on average per embryo is 171,551,080, 103,472,660 and 161,117,565 for genomic, RNA and WGBS experiments, respectively (supplementary Table S1).

Methylome analyses

WGBS data

An average of 161.1 gigabases (Gb) of sequence were generated per embryo from whole genome bisulfite sequencing. After removing adaptor sequences and lower quality reads, about 109.6 Gb clean reads were mapped with an average of 70.4% of uniquely mapped paired-end reads (supplementary Table S2). An average of 94.0 Gb deduplicated reads were kept to analyse DNA methylation (supplementary Table S2, supplementary Figure S1). The bisulfite conversion rate was greater than 99% for each sample (mean 99.3%). Methylation profiles were observed in three different sequence contexts : CG (CpG dinucleotides) and non-CG contexts (CHH and CHG where H is A, T or C). As the CpG methylation mediated by DNMT1 (DNA methyltransferase 1) leads to symmetrical methylation between both DNA strands (see¹²), the methylation information on both strands was added at each detected CpG site. We observed an average of 12.5 millions of CG positions (average depth - 9.2X) which represent 69.7% of the CG from the reference (17888534 CG positions in the chicken genome version 5 from Ensembl (<http://www.ensembl.org/info/data/ftp/index.html>)). This number is quite lower than that observed in other WGBS studies in chicken (near 85% in Li et al¹³, 95% in Zhang et al¹⁴). This may be due to the version of the genome used, and the sequencing depth. Global methylation level was about 66.7% in CpG context, whereas the methylation rate was lower in the non-CG contexts, with methylation rates reaching 0.7% and 0.8% for CHG and CHH sites, respectively. These methylation levels are slightly higher than the levels observed in other studies on muscle (Zhang et al¹⁵ : 63.2%, 0.4% and 0.4% methylation respectively) of juvenile and adult, or blood (Zhang et al¹⁴ : 51.22%, 0.4% and 0.45% methylation respectively) of adult chicken. These differences may be explained by the tissue used - here we analysed whole embryos, that is a combination of different tissues - and most probably by the age of the individuals: it has long been known that global methylation decreases with aging, at least in human^{16,17,18}. Thanks to the genomic sequences available for each embryo, 9.7 millions of SNP positions were detected and removed from methylome data (11.8% of detected CG). Due to the low coverage of our WGBS data, we set a threshold of 5X as a minimum depth for 75% of embryos, i.e. for all common positions across seven embryos, five of them had a minimum depth of 5x. Thus 5982957 common CG sites across the samples (average depth 12.2X) were analysed, representing about 33.5% of the CpG sites present in the chicken genome (supplementary Table S3).

Methylation profiles in different functional regions

Methylation levels fluctuate across the genome according to different genomic and functional regions¹⁹. We thus analyzed the methylation landscape of chicken embryos along with their genomic position and functional features. From the chicken genome annotation (Ensembl release 89) we defined a promoter as 1 kb upstream to the transcription start site (TSS). Each detected CG was annotated in one feature. Overall, mean methylation is lower in promoters (46.4%) than in genes (71.6%) (Figure 1). The proportion of CGs with a low methylation level is higher in promoters as compared with gene bodies, while more CGs are fully methylated in gene bodies than in promoters. Across the genic features, the methylation level was the lowest in 5'UTR (5' untranslated transcribed region) (44.0%) whereas a high methylation rate was observed in coding DNA sequences (77.5%), in introns (70.8%) and in 3'UTR (72.5%) (Figure 2). These results are in accordance with previous methylation levels observed in chicken^{13,20}.

Methylation of CpG islands (CGI) is usually associated with gene regulation²¹. We defined CpG islands from the Ensembl 89 annotation as presenting a GC content greater than 50%, a ratio of observed CpG / expected CpG greater than 0.6, a minimum of 200 bp of length. With these criteria from EMBOSS version 6.4.0.0, we obtained a total of 59116 CGI. Shore regions were defined as the regions immediately flanking the CGI (+/- 1 kb), and shelf regions as 1 kb regions flanking the shores (from 1 kb to 2 kb from the CGI). The other regions of the genome were considered as openSea. The mean methylation level is lower in CGI (45.3%) as compared with neighbouring shores, shelves and openSeas: their methylation levels reach 54.7% for shores, 69.8% for shelves and 73.8% for openSea (Figure 3).

The methylation level is lower in CGI than in flanking regions whatever the functional localisation of the CGI (Figure 4). Moreover, the CGI methylation level is lower in promoter regions than in genes and intergenic regions. The bell-shape

observed in the "genes" and "intergenic" profile are probably due to the decrease in CG methylation present near the promoter (Figure 5), because genic and intergenic CGI have one (genic) or at least one (intergenic) of their extremities flanking a promoter.

Repetitive elements are usually methylated to maintain genomic integrity and stability²². Our results showed that the methylation level is about 75.1% in repeat elements (Ensembl 89) and 72.1% in their 2kb flanking regions in all samples (supplementary Figure S2). While in mammals repeat elements are mostly hypermethylated¹², these result differ from the observations of Li et al and Zhang et al^{13,20}, who showed a lack of hypermethylation in chicken muscle and lung. On the contrary our results confirm the first global analyses of repetitive DNA methylation in chicken²³. Further studies should explore the dynamic of DNA methylation in chicken during development, including detailed analyses of different families of repetitive elements, and putative tissue-specific mechanisms.

CTCF (CCCTC binding factor) plays an important role in genomic interactions, gene expression and in the regulation of chromatin architecture²⁴. CTCF patterns in chicken were determined by Fr-AgENCODE²⁵. In our results, the DNA methylation is low in CTCF sites (40.1%), and wave-shaped in CTCF flanking regions. These oscillations of methylation levels in flanking regions around CTCF are characteristic of the presence of nucleosomes (open and condensed chromatin with unmethylated and methylated cytosines respectively) (supplementary Figure S3). This particular pattern has been observed in other vertebrates, as in mouse²⁶ and human²⁷.

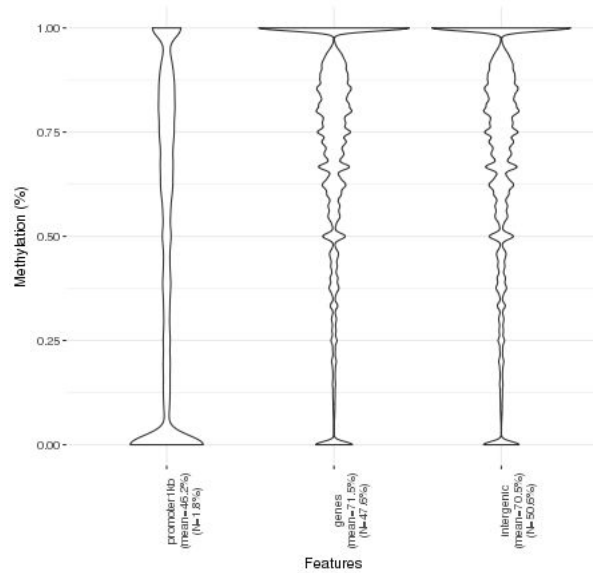


Figure 1. Methylation rate in genomic features (genic, promoter and intergenic regions)

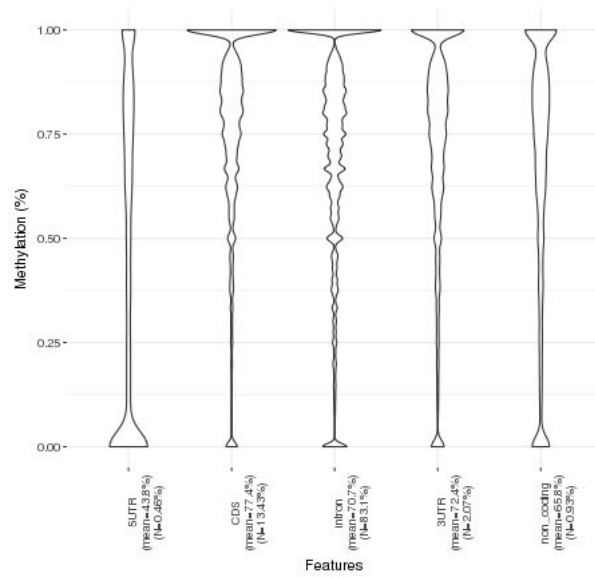


Figure 2. Methylation rate within genic features (5'UTR, CDS, intron, 3'UTR and non coding regions).

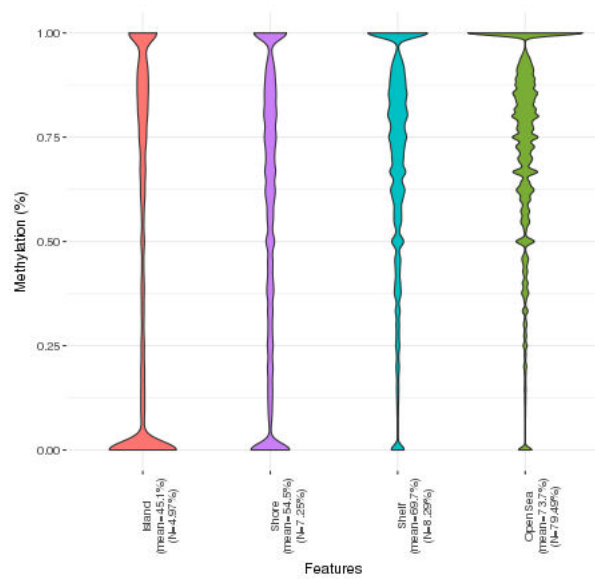


Figure 3. Methylation level in CpG islands and around.

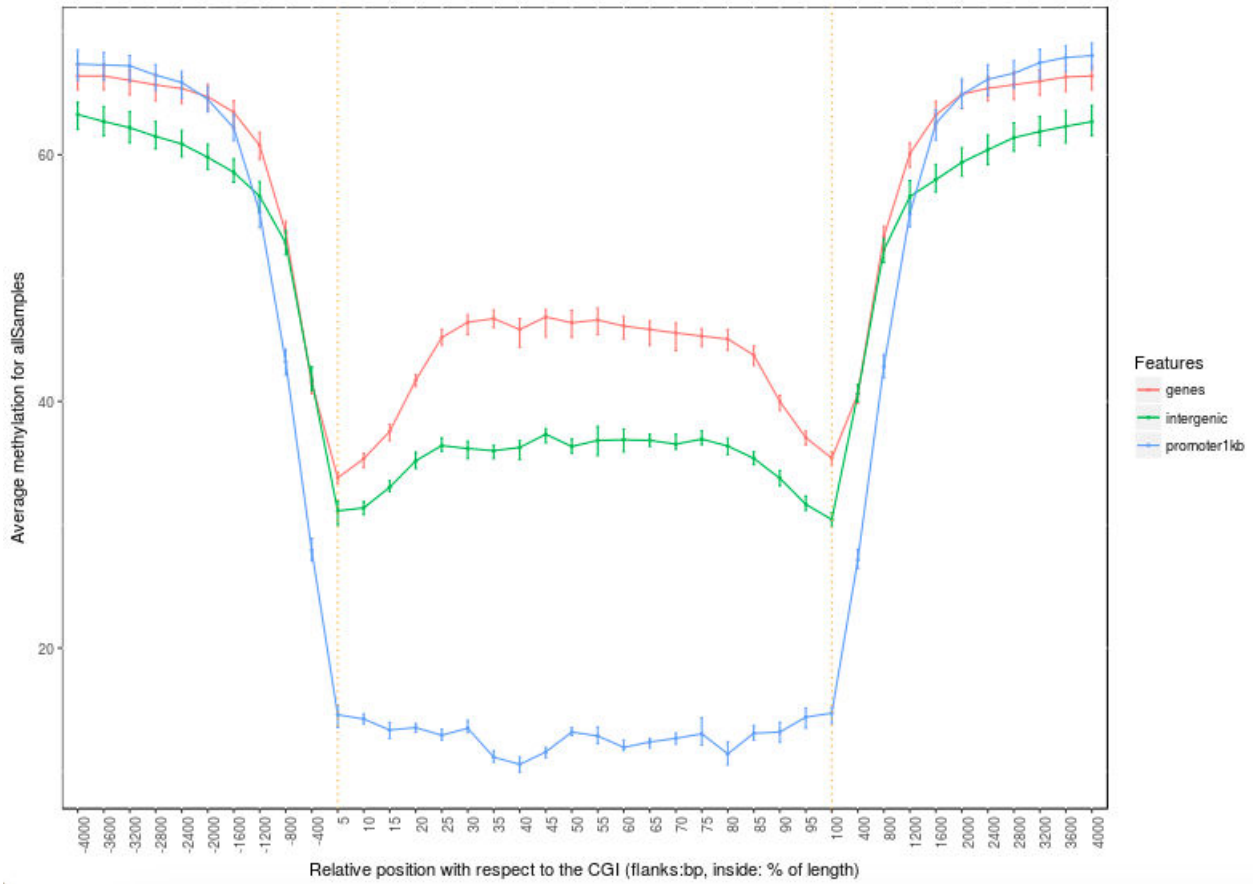


Figure 4. Distribution of the mean methylation level across the genome.

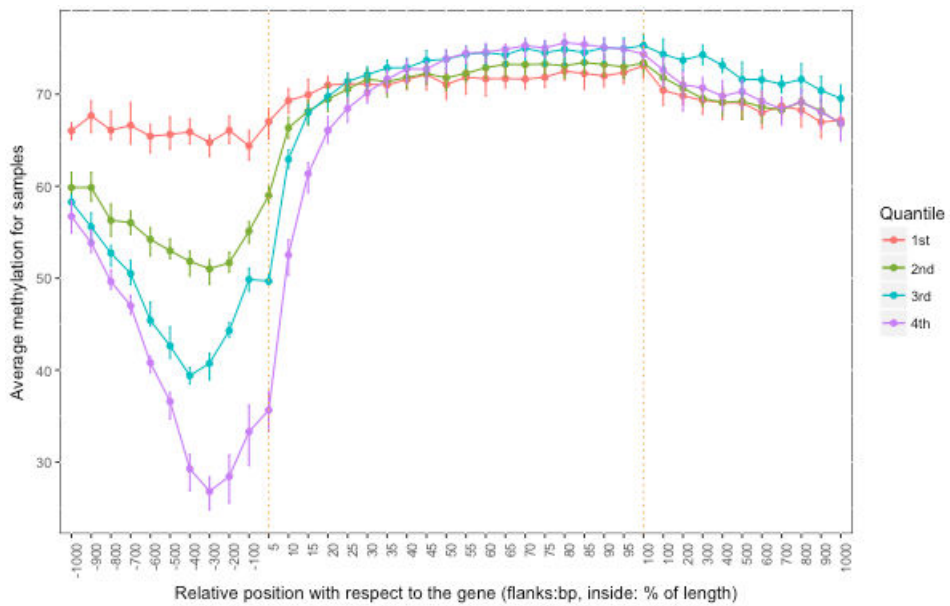


Figure 5. Distribution of the mean methylation level associated with 4 quantiles of expression. The methylation profile in gene body was drawn as the percent of the length for each gene.

Relation between DNA methylation and gene expression

The methylome and transcriptome data available for each embryo were used to observe the link between expression and methylation for each gene. Based on the expression levels, all genes were divided into four quantiles, from the lowest 25% and to the highest 25% (supplementary Figure S4). In the first global analysis between the methylation profile in promoter (1 kb upstream of the TSS) and the expression levels, we found a significantly decrease in the methylation rate according to the quantiles (Figure 6). The methylation profiles were also analysed across the gene body, and 1kb after the end of the gene. The profile of the methylation level is clearly specific for each expression quantile in the promoter region, with the lower the promoter methylation level, the higher the gene expression (Figure 5). The methylation level then increases downstream of the TSS, with differences between expression quantiles nearly absent in the gene body and the 3' non coding region. This pattern has already been highlighted, in several tissues and different species, as chicken or human^{13,28,29}.

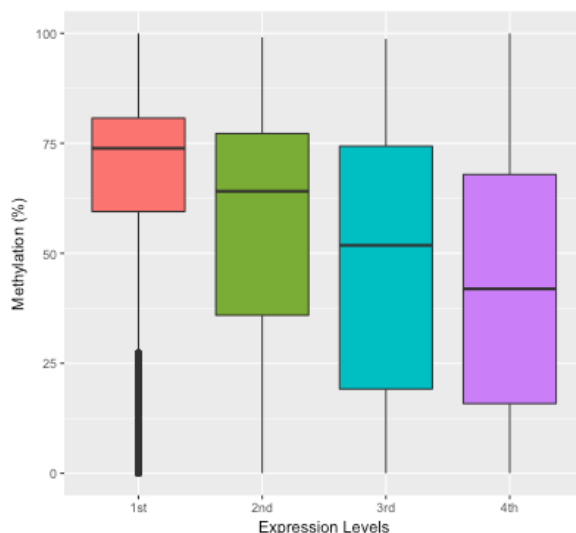


Figure 6. Distribution of methylation level in each quantile of expression.

Differential genome regulation

Differential methylation

We then observed if there were differences in methylation profiles between sexes or between cross ways. The overall methylation pattern is very similar between sexes on the autosomes, and show slight differences on the Z chromosome (supplementary Figure S5). The chicken methylation profile on male chromosome Z (supplementary Figure S5) lacks the clear hemimethylated pattern associated with X-inactivation in mammals³⁰. We detected specific positions with different methylation levels between female (n=4) and male (n=3) embryos. A total of 36576 DMC (differentially methylated cytosines) were detected (supplementary Table S4). Of these DMC, 24162 were hypermethylated in males (66%), including the MHM region (Male HyperMethylated³¹) (supplementary Figure S6).

These DMC are mainly localised in intergenic regions (53.5%). From these DMC, 970 DMR (differentially methylated regions) were detected (supplementary Table S5). The DMR mean length was 146.9 bp, containing 6.5 CGs on average. The longer DMR (5103 bp) was found on chromosome Z, overlapping a long non-coding RNA (ENSGALG00000044868), with unknown function. Four clusters of short DMRs could be found on chromosomes 7 (positions 31597 - 33856), 13 (18368120 - 18372874), 15 (13105 - 17885), and 27 (5629029 - 5635581), all in intergenic regions. A total of 427 annotated genes were associated with the detected DMRs. A GO overrepresentation test (FDR < 0.05) highlighted two enriched biological processes, MAPK (mitogen-activated protein kinase) cascade and regulation of catalytic activity. MAP kinases are involved in a wide range of biological functions, and have been shown to be required in sex development in mammals^{32,33}.

Between the two reciprocal crosses, we detected 26615 DMC with the same distribution as the differential regions between females and males (53.5% in intergenic regions and 44.8% in genic regions). A total of 852 DMR were detected, corresponding to 370 unique genes. Four biological processes were significantly enriched: cell adhesion, single-multicellular organism process, developmental process and cell communication. These processes are involved in embryonic development, and no clear specific pathway could be highlighted to differentiate both cross ways. It would be interesting to measure different phenotypes

on adults from the same crosses, to link differential methylation and expression during development with putative differential traits later in life.

From all the detected DMC, less than 1% (n=622) were common between sex-DMCs and cross-DMCs, confirming the specificity of the significant results observed. The genomic repartition of the DMC was similar in both analyses, with less cross-DMC than sex-DMC in each chromosome, even on the Z one (supplementary Figure S7).

Paragraph on genomic imprinting to be written.

Link between differential methylation and differential expression

Differential expression analyses were performed on a subset of 14501 genes for which at least 2 samples have more than one count per million of reads. A total of 175 genes were differentially expressed between males and females (FDR < 0.05, supplementary Table S6), corresponding to 1.2% of the analysed genes. Of these, 34 mapped to autosomes (and 47 to scaffolds), with 26 genes being up-regulated and 8 down-regulated in females. GO enrichment analysis didn't show any significant enrichment except in hemoglobin complex (GO cellular component, $p = 1.72E-07$). Differences in blood compositions are well-known between males and females^{34,35}. All the 94 differentially expressed gene mapped to chromosome Z were upregulated in males. The low number of autosomal genes differentially expressed between males and females may be explained by the fact that we analysed the whole embryo instead of specific tissues and, most importantly, that the gonads are not differentiated at this stage; the extent of sex-biased expression actually increases as a function of age³⁶. Ayers et al³⁷ observed more than twice our numbers of DGE from chicken embryos at the same stage (4.5 days), for a similar dataset per sample, but they focused their analysis to gonads.

Only one gene was found differentially expressed with a promoter differentially methylated between sexes : H1F0. This gene appears as more expressed in females than in males ($\log_{2}FC = -9.36$) and the associated DMC is less methylated in female (difference methylation = -16.32%). It has been shown recently³⁸ that H1.10 levels are lower in male than in female primordial germ cells (PGC) in mouse embryos at E12.5, a developmental stage comparable to chicken E4.5. H1.10 is a replication-independent subtype of histone H1, as is H1.0, the protein encoded by the H1F0 gene. Due to the analysis of whole chicken embryos in this study, the observed difference is probably not fully explained by differences in PGC, but it is of interest to note the consistency of these results: a gender-specific quantity of H1 subtypes may come with specific chromatin structures, more or less permissive for epigenetic reprogramming.

For the DMC localized in gene body, 28 differential genes were found between sexes (supplementary Table S7), 7 of them being localized on autosomes. These genes represent different cellular processes. Notably, 2 genes (HBBR and SPTB) are related to blood, as observed above.

Comparing differentially expressed genes and differentially methylated regions led to the identification of two additional genes, both overexpressed in males and with an intronic DMR hypermethylated in males: COL6A3 (Collagen Type VI Alpha 3 Chain), and KRT14 (keratin 14). We cannot conclude from our data that the DNA methylation governs the regulation of the expression of the associated gene, but these results are in accordance with the positive regulation of expression by intronic methylation already observed³⁹.

Due to the very low number of genes differentially expressed between cross ways, only one gene showed differential expression and methylation (A1CF or APOBEC1 complementation factor). This gene is more expressed in the female 6 * male R- cross than in the reciprocal cross, and the methylation level in the two related DMCs is lower in female 6 * male R- cross. These DMCs are localised in an intronic region of the gene. If they are involved in the regulation of A1CF transcription, this could be determined by a methylated-dependant intronic enhancer⁴⁰. The ACF1 gene is involved in C-to-U RNA editing as a complement of the APOBEC1 enzyme, but, while A-to-G editing does exist in chicken, the C-to-U phenomenon seems to be absent in this species⁴¹. The physiological role of A1CF still needs to be characterized⁴².

Allele-specific methylation

This part is in progress and some details will be added from the thesis manuscript.

Our animal designed was set to maximise the number and informativity of SNPs in the F1 embryos⁷. Moreover, the genomic sequences of the F0 parents were used to build a diploid genome as a reference. Thus, we analysed allele-specific methylation, in relation with allele-specific expression. For each cross way (7127 or 8303), reads from each embryo were aligned against the corresponding diploid genome. From the 161.1 Gb raw data, about 52.8 Gb clean read were mapped with an average of

33.9% of uniquely mapped paired-end reads. Then about 45.3 Gb deduplicated reads were generated (14.5% of removed reads) to analyse ASM (supplementary Table S8). About 13.3 millions of CG positions (depth - 4.9X) were detected, representing 37.2% of the CG from the diploid reference. The number of observed CG positions is lower than the number observed in the first analysis (mapping against the reference): due to the diploid genome, the mean depth is about half the previous one, and the required minimum coverage filters far more positions. Global methylation level was about 71.6% in CpG context (supplementary Table S9).

The ASM analyses were performed with the same pipeline used for the non allelic analyses, with unchanged parameters. We obtained a list of CG observed from the maternal and the paternal genomes per individual, with their respective methylation levels. A total of 155022 CpG filtered positions were common between all the embryos. From these, 4069 (2.6%) showed differential methylation between both alleles. These results come from whole embryos and, as an average, may underestimate the actual number of ASM sites. Nevertheless, this is in accordance with several ASM studies in human, detecting from 1.5% to 22% of CpG sites showing a genetic cis-regulation (see⁴³). These results suggest a genetic regulation of the methylation of several CpG sites in chicken. At a time when GWAS (Genome Wide Association Study) will be more and more complemented by EWAS (Epigenome Wide Association Study) analyses⁴³, the existence of cis-regulation of epigenetic variability has to be kept in mind to avoid wrongly declaring non-genetic associations.

The ASE analyses gave a total of 1412 ASE genes (supplementary Table S10). This proportion of ASE genes (9.7% of expressed genes) is lower than the results obtained by Zhuo et al⁴⁴ who found between 17.3% and 18.3% of genes showing ASE patterns in embryonic brain in liver. It is at least partly explained by the fact that we analyzed whole embryos, while more than half of ASE genes seem to be tissue specific⁴⁴.

The few number of sites showing both ASM and ASE is partly due to the sequencing depth used for the WGBS study: mapping to a personal diploid genome more than halves the mean depth, and the total observable CpG sites was reduced. The number of associations ASM/ASE would certainly be higher with a greater depth, even if it has been suggested that ASM does not seem to be strongly involved in the regulation of gene expression⁴⁵. Allele specific expression may be the result of several mechanisms, not always involving ASM, as transcription factor binding sites affected by a SNP⁴⁶.

Incomplete dosage compensation

To observe the global behaviour of autosomes and Z chromosomes between sexes, we only kept the genes mapped to known autosomes or Z chromosome, with $TPM \geq 1$ for every sample, ending with 11,892 genes (11,318 from autosomes and 574 from Z chromosome). The average expression on Z chromosomes, expressed as $\log_2(TPM)$, was significantly lower in females than in males (ratio = 1.18, p-value = 2.245E-11, supplementary Figure S8). The ratio in median expression levels between Z and autosomes ranged from 0.86 to 0.88 in females and from 0.97 to 0.99 in males. As already known^{47,48}, we confirm that dosage compensation, that is compensation in expression of the different copy numbers of sex chromosomes, is not complete in chicken. The Z-to-A ratio observed here between males (ZZ, averaged ratio = 0.98) and females (ZW, averaged ratio = 0.87) confirm the results obtained by others, while being slightly higher^{44,47}.

Conclusion

To be completed

The chicken whole-embryo methylome described here confirms the overall characteristics observed in vertebrates: most of the CpG sites are highly methylated, CpG islands and promoters have a lower methylation level, promoter methylation is negatively correlated with gene expression. We show that allele-specific methylation does occur in chicken, at a level comparable to other organisms. The association between ASM and ASE is not frequent, reflecting probably a biological process, ASE being also the result of several other biological mechanisms. This few occurrences could also be the result of our experimental design, because the samples used are a mix of several tissues, and the expression and methylation levels observed are an average, at the whole embryo scale, of different mechanisms probably occurring in different tissues. Our data also confirm the absence of genomic imprinting at the methylation level.

More and more studies are being set up to decipher the genetic and environmental causes of phenotypic variability in farm animals, the latter being notably conveyed by epigenetic mechanisms. These analyses have to take into account the genetic variability of epigenetic marks. Our study is a first step towards the analysis of the cis-regulation of CpG methylation by neighbouring SNPs.

Methods

Ethics statement

Chickens were bred in accordance with European Union Guidelines for animal care, following the Council Directives 98/58/EC and 86/609/EEC at INRA, UE1295 Pôle d'Expérimentation Avicole de Tours, F-37380 Nouzilly. Animals were maintained under standard breeding conditions. The farm is registered by the Ministry of Agriculture with the license number C37-175-1 for animal experimentation. The experiment was realized under authorization 37-002 delivered to D. Gourichon.

Tissues dataset

The material used in this study for the embryo sequences dataset was previously described⁷. Briefly, two chicken lines were crossed, Line 6⁴⁹ and Line R-⁵⁰. Seven embryos were produced from 2 families. Embryos were kept as a whole. Genomic DNA and total RNA were concurrently extracted from the same samples of whole embryos (AllPrep DNA/RNA Mini Kit, Qiagen). RNA quality was measured by a BioAnalyzer (Agilent); all samples had a RIN (RNA Integrity Number) ≥ 9.9 .

Sequencing

RNA and DNA sequencing

The RNA and DNA datasets have been previously described⁷. We used here a subset of these data, for which the WGBS data were available. Briefly, a total of 7 embryo RNA libraries were sequenced (paired-ends, 100 bp) in triplicate after multiplexing, by randomizing their position in 6 different sequencing lanes and their DNA was sequenced (paired-end, 100bp) on an Illumina HiSeq 2000.

Whole Genome Bisulfite sequencing

From the DNA used for genomic sequencing, WGBS library preparations were performed through the NextFlex bisulfite sequencing kit (BIOO Scientific Corp) following the manufacturer's instructions. The obtained libraries were sequenced through paired-ends, 2x100bp, on an Illumina HiSeq 2000 sequencer (TrueSeq V3 kits). To check for complete bisulfite conversion, 2 ng of control unmethylated pUC19 DNA (Zymo Research) was spiked-in.

Computational analyses

Genomic sequences analyses

Sequences were aligned to the chicken genome assembly (Gallus_gallus-5.0,⁵¹ using the BWA program version 0.7.0, option mem⁵². SNP (Single Nucleotide Polymorphism) calling was performed using the Genome Analysis Toolkit (GATK, version 3.5) software package⁵³ with Picard tools (<http://broadinstitute.github.io/picard>) as PCR duplicates detector and GATK UnifiedGenotyper as SNP caller. SNPs were filtered through GATK variant quality score recalibration and SnpSift⁵⁴ for variation quality ≥ 30 , depth ≥ 10 and keeping only biallelic SNPs.

RNA sequences analyses

The read qualities of the RNA-Seq libraries were evaluated using FastQC⁵⁵. The analyses were performed using a standard workflow⁵⁶: we used the STAR aligner to map reads against the chicken genome assembly Gallus_gallus-5.0⁵⁷, and RSEM to quantify the expression levels of genes and transcripts⁵⁸. RSEM expected counts were used for differential analyses with edgeR⁵⁹, for the genes with FPKM > 1 in at least 2 samples. Gene Ontology (GO) enrichments were performed through the Gene Ontology Consortium tool (<http://www.geneontology.org/page/go-enrichment-analysis>), based on the PANTHER classification system⁶⁰.

WGBS analyses

Methylation analyses were performed through an in-house pipeline using standard tools. Raw reads produced by Illumina HiSeq were trimmed for adapter sequences using trim_galore v0.4.0 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and shorter sequences were removed (< 40 bp). A filter from sequencing based on a quality score (> 20 : probability of 1 incorrect base in 100) was applied using FastQC⁵⁵. These trimmed reads were aligned to the fifth version of the chicken genome (Warren2017) using Bismark version 0.16.1⁶¹ with Bowtie2-2.2.9⁶² to identify methylated and unmethylated cytosines. After removing PCR duplicates with SAMtools version 1.3.1⁶³, the methylation rate for each CG dinucleotide was defined as the ratio of the number of methylated cytosines to the total of methylated and unmethylated cytosines at a given site using the MethylKit R package, version 0.9.4⁶⁴. The detected positions were normalized by median coverage to remove experimental variation across samples. Only sites covered by a minimum of 5 reads in 75% of samples were kept in order to maximize the relevance of the observed methylation level. SNP positions detected from the genomic data were removed from the CG dinucleotide list to avoid bias of the methylation rate. Only common CG positions across samples were kept. To detect the differential methylated cytosines (DMC) and regions (DMR) between sexes and cross ways, we used the DSS R package, version 2.14.0⁶⁵. A dispersion shrinkage method for estimating the dispersion parameter

from Beta-Binomial distributions was used⁶⁵. Studies showed that methylation levels influence the neighboring CpG sites⁶⁶. Thus a smoothing parameter was activated to better estimate mean methylation at each site, with a smoothing span of 120 bp. An FDR (False Discovery Rate) of 0.05 was used to accept a DMC as significant. Then a DMR was defined as containing a minimum of three CpG, 33% of CG sites with significant p-values, a difference of 10% between the methylation levels of the two groups and a threshold of p-values of 5%⁶⁷. The normalized methylation data were used for a comparison with expression levels, expressed as TPM (Transcripts Per Million).

Allele-specific analyses

To perform allele-specific analyses, we chose to discriminate each expressed or methylated allele through allele-specific mapping: we first built personal diploid genomes based on *Gallus_gallus*-5.0 by including all SNPs observed from the parental DNA sequences into 4 fasta files, one for each parent, by using the `vcf-consensus` option from `vcftools`, version 0.1.12a⁶⁸. Parental genomes were then concatenated, each of the resulting fasta file containing two versions (maternal and paternal) of each chromosome. A "personal diploid genome" was thus available for each cross way. For both allele-specific expression (ASE) and allele-specific methylation (ASM), we performed the same analyses as described above, replacing the reference genome by the personal diploid genome. Differential expression and differential methylation were then calculated between the paternal and the maternal genomes.

Acknowledgements

We thank the entire staff of the PEAT experimental unit for their excellent animal care. We thank Marianne DeGorter for proofreading of the final version. Sequencing was performed at GeT-PlaGe Genotoul platform (<http://get.genotoul.fr/en/>). We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and/or computing and/or storage resources. This study was supported by the French "Agence Nationale de la Recherche" EpiBird [ANR-009-GENM-004] and ChickStress [ANR-13-ADAP-0014] grants, and the INRA Animal Genetics Division.

Supplementary information

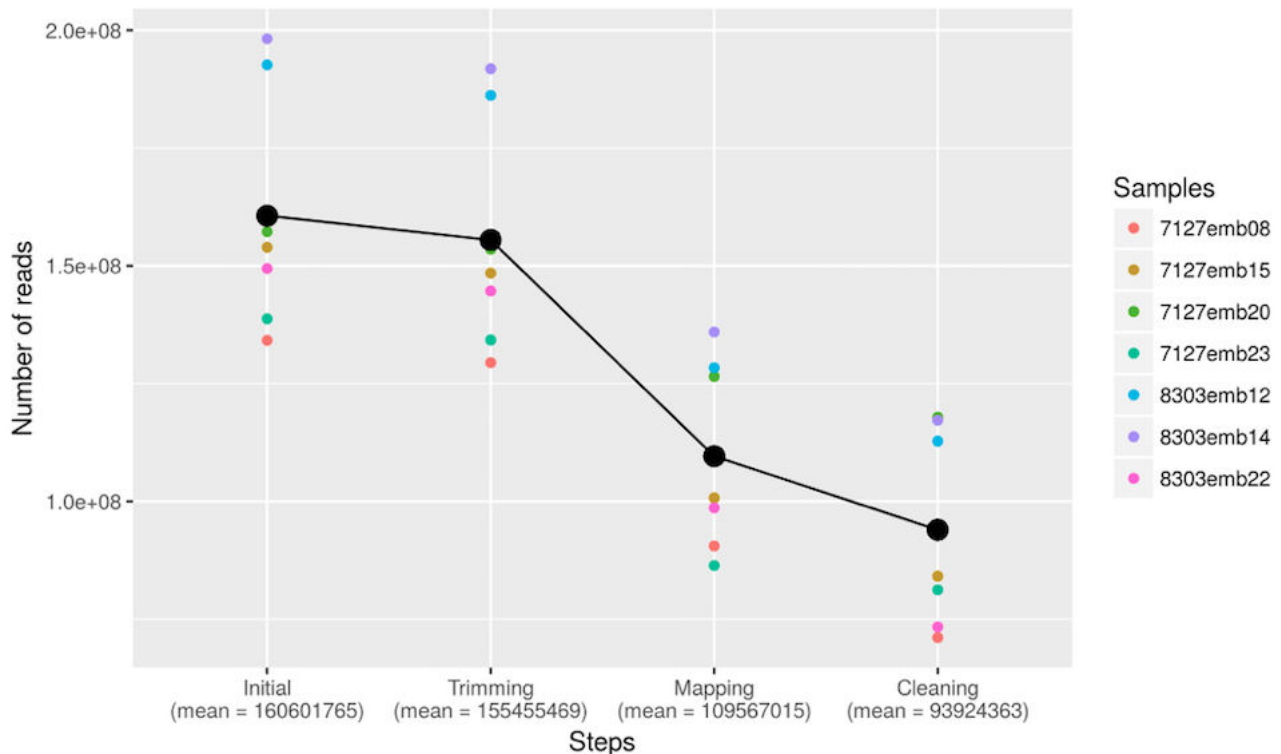


Figure S1. Distribution of reads after each pipeline step.

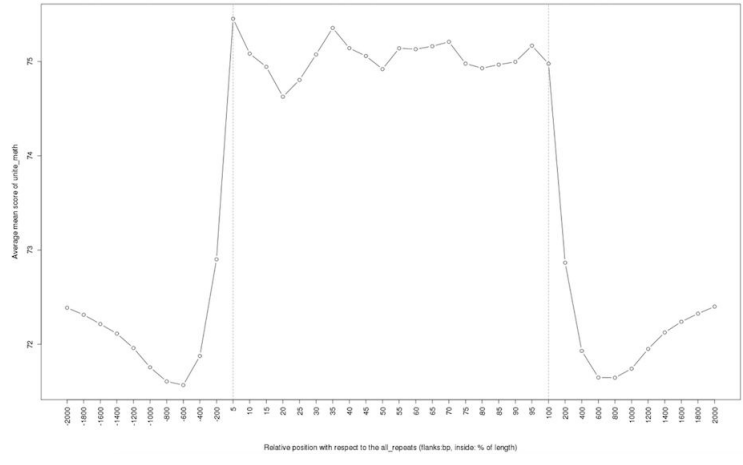


Figure S2. Methylation level in repeat regions.

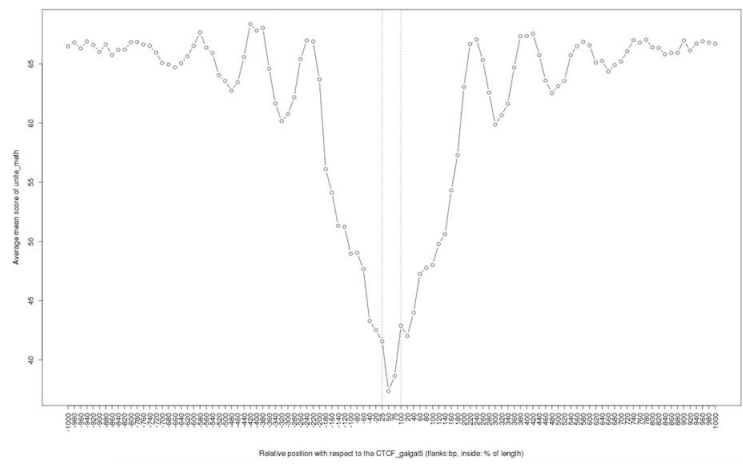


Figure S3. Methylation level in CTCF binding sites.

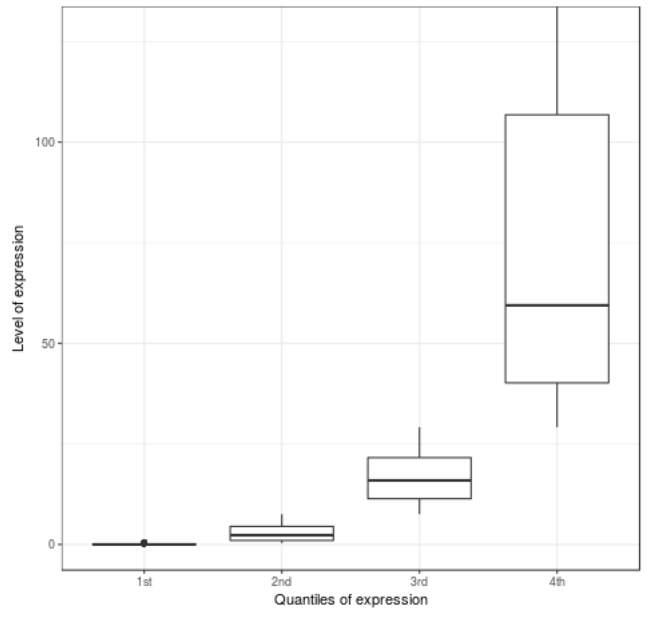


Figure S4. Distribution of quantiles of expression.

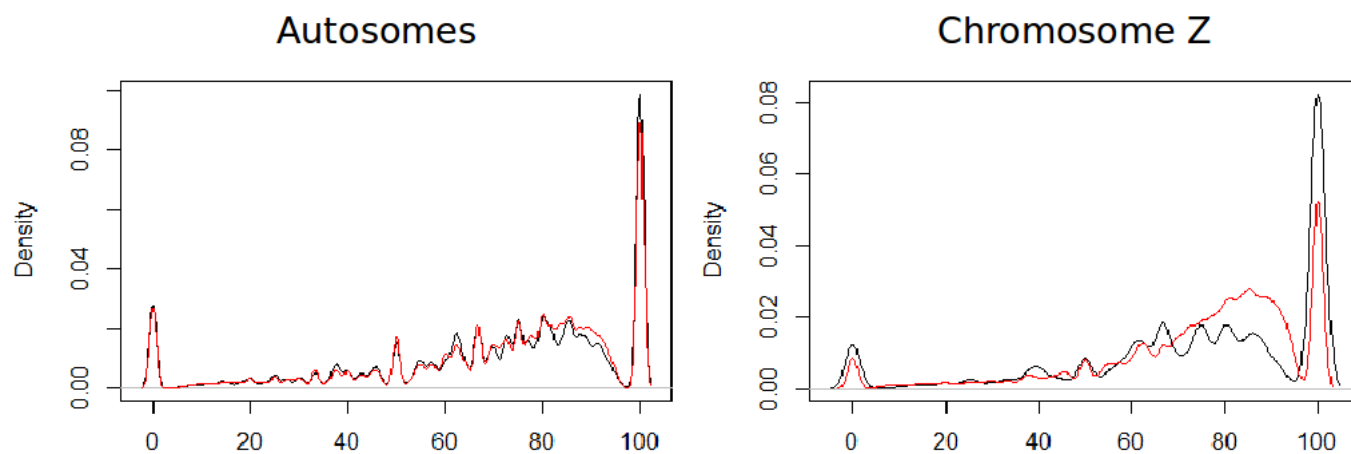


Figure S5. Distribution (kernel density estimates) of methylation levels by sex. Left: autosomes, right: chromosome Z; red: males, black: females.

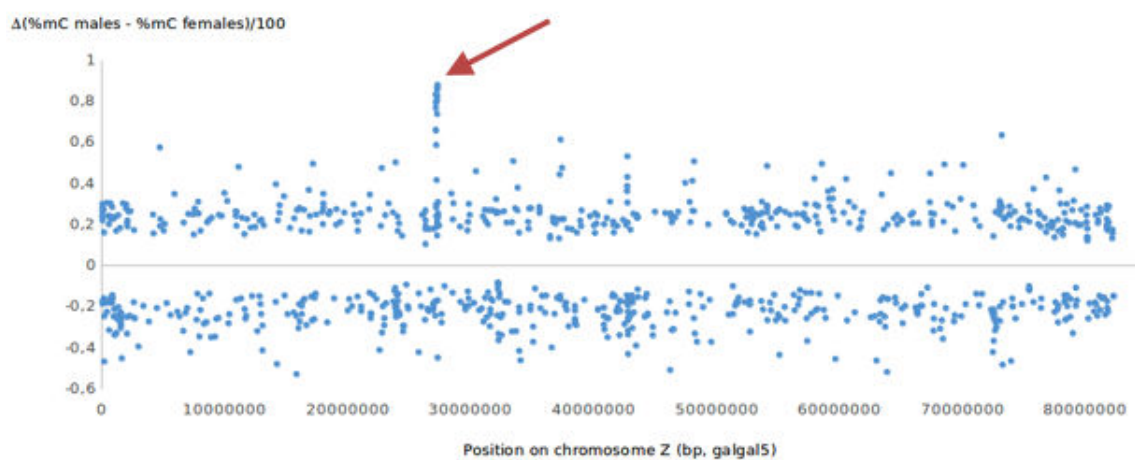


Figure S6. DMC observed between males and females on chromosome Z. The y-axis shows differences in mean methylation levels between males and females, along chromosome Z (x-axis). Each point represents a DMC, with a positive value for DMC hypermethylated in males compared to females. MHM (Male Hypermethylated) region is shown (red array).

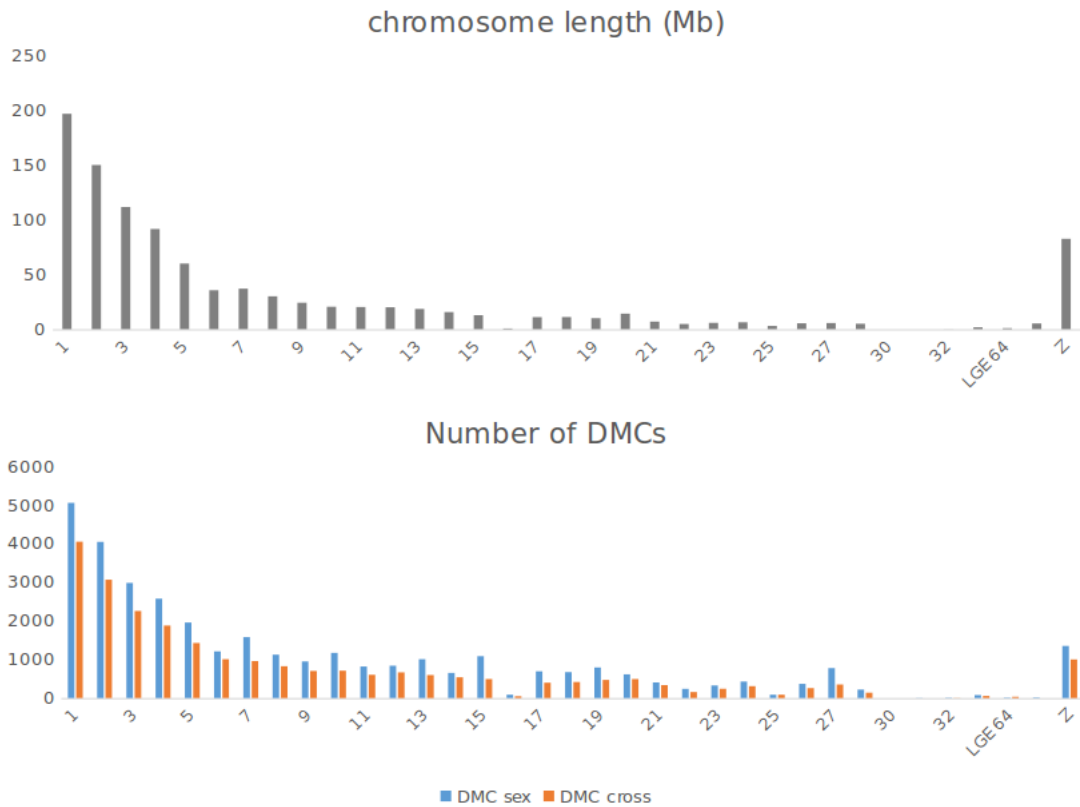


Figure S7. Number of DMC observed per chromosome. The x-axis shows the chromosome names. Upper panel: chromosome length (Mb). Lower panel: number of observed DMC (blue: between sexes; orange: between cross ways).

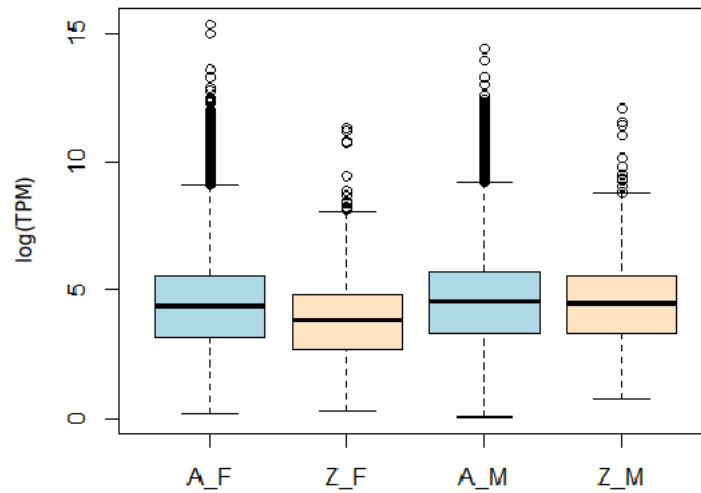


Figure S8. Comparison of gene expression according to sex and chromosome category. Left: females, right: males. Autosomes (11,318 genes): light blue, Z chromosome (574 genes): beige. The average expression on Z chromosomes was significantly lower in females than in males (p -value = $2.567E-11$, Wilcoxon rank sum test).

Sample	SRA_Biosample_ID	DNaseSeq	RNASeq	WGBS	Cross	Sex
7127emb08	SRS606787	160931428	96252590	134142902	7127	F
7127emb15	SRS607166	158108638	162702222	153903838	7127	M
7127emb20	SRS607168	180478796	159083960	157230624	7127	M
7127emb23	SRS607177	174648166	107254648	138736708	7127	F
8303emb12	SRS607183	154397208	66372756	192625836	8303	F
8303emb14	SRS607184	164040856	87966494	198175730	8303	M
8303emb22	SRS607186	170346290	68295294	149396714	8303	F

Table S1. Table of the number of reads for genome, transcriptome and methylome experiments.

Samples	Raw reads	Reads after mapping	Mapping rate (%)	Reads after cleaning	Duplication rate (%)
7127emb08	134142902	90516586	69.90	71061394	21.49
7127emb15	153903838	100721834	67.83	84077536	16.53
7127emb20	157230624	126741712	82.60	117979692	6.91
7127emb23	138736708	86366098	64.30	81204324	5.98
8303emb12	192625836	128346674	68.93	112779862	12.13
8303emb14	198175730	135945424	70.87	117196102	13.79
8303emb22	149396714	98614386	68.17	73338668	25.63

Table S2. Description of the WGBS data for each embryo.

Samples	Detected CG	Depth	mCpG percent (%)	Depth of common CG positions
7127emb08	12424455	7.01	65.45	12.07
7127emb15	12012435	8.79	69.71	13.64
7127emb20	12117282	8.99	66.21	10.96
7127emb23	12573095	9.15	66.43	12.00
8303emb12	13061221	11.32	66.07	12.28
8303emb14	12852618	11.80	65.95	12.64
8303emb22	12175001	7.46	67.35	11.99

Table S3. Description of the cytosine positions in CpG context and the global methylation analysis for each embryo.

Table S4. Differentially methylated cytosines between males and females. This long table is here :
http://genoweb.toulouse.inra.fr/~mmersch/Article_Mersch-et-al/suppDMCsex.pdf

Table S5. Differentially methylated regions between males and females (*defined as the sum of the test statistics of all CpG sites within the DMR). This long table is here :
http://genoweb.toulouse.inra.fr/~mmersch/Article_Mersch-et-al/suppDMRsex.pdf

Table S6. Differential expressed genes between males and females (logFC > 0 : males more expressed).

ID	Gene name	Chromosome	Start	End	logFC	logCPM	PValue	FDR
ENSGALG00000044240	H1FO	1	51163992	51164879	-1,38	5,99	2,70E-07	5,00E-05
ENSGALG00000015079	UPK1B	1	80776274	80787066	1,16	5,32	1,00E-04	9,90E-03
ENSGALG00000044745	ENSGALG00000044745	1	160124743	160139699	-1,08	5,79	6,60E-06	9,10E-04
ENSGALG00000017137	GJA3	1	179008069	179009604	-2,93	2,29	2,30E-05	2,90E-03
ENSGALG00000017347	HBE	1	194562710	194576539	-1,13	13,22	6,90E-05	7,00E-03
ENSGALG00000006189	CRYGN	2	6282029	6286517	-3,07	0,74	7,50E-05	7,40E-03
ENSGALG00000034585	CP49	2	42762444	42777858	-2,78	3,6	6,70E-07	1,10E-04
ENSGALG00000033428	ENSGALG00000033428	2	123250307	123260107	-1,3	4,91	1,00E-04	9,90E-03
ENSGALG00000008735	BFSP1	3	11489065	11507709	-3,27	5,47	2,90E-08	6,30E-06
ENSGALG00000008815	LRRN4	3	15330848	15337331	1,21	5,72	3,80E-05	4,50E-03
ENSGALG00000011096	ABCB10	3	40180859	40197033	-0,73	5,71	4,60E-04	3,50E-02
ENSGALG00000016498	MFSD2B	3	104544323	104575302	-1,43	3,34	7,10E-04	5,00E-02
ENSGALG00000012298	SLC39A8	4	61121025	61141216	-1,21	4,51	2,80E-04	2,30E-02
ENSGALG00000036805	SPTB	5	311493	328512	-1,06	6,67	2,50E-04	2,10E-02
ENSGALG00000009552	ENSGALG00000009552	5	28954862	28959674	-1,4	4,37	4,50E-05	5,10E-03
ENSGALG00000003923	COL6A3	7	4808259	4861524	0,95	7,21	4,80E-04	3,70E-02
ENSGALG00000032039	HOXD13	7	16389450	16392023	3,12	3,58	1,10E-04	1,00E-02
ENSGALG00000010219	ENSGALG00000010219	8	21007876	21011979	-0,86	5,75	2,20E-04	1,90E-02
ENSGALG00000006520	MYH11	14	7610019	7666669	1,31	8,3	2,60E-04	2,20E-02
ENSGALG00000023740	HBZ	14	12156905	12158534	-1,18	13,16	6,80E-05	6,90E-03
ENSGALG00000031597	HBAD	14	12160597	12161521	-1,18	10,68	5,40E-05	5,90E-03
ENSGALG00000043234	HBAA	14	12163617	12164434	-1,14	11,58	4,90E-04	3,70E-02
ENSGALG00000005630	CRYBB1	15	7224895	7227195	-2,55	2,08	1,20E-05	1,50E-03
ENSGALG00000005634	CRYBA4	15	7229327	7230473	-2,94	0,58	1,10E-04	1,00E-02
ENSGALG00000002576	ASL2	19	4919234	4928185	-2,09	5,67	3,70E-05	4,40E-03
ENSGALG00000021230	EPB42	20	6889947	6899216	-1,28	5,35	2,60E-04	2,10E-02
ENSGALG00000007447	TUBB1	20	11089454	11094495	-1,08	6,84	6,70E-04	4,80E-02
ENSGALG00000001407	ADAMTS15	24	1676278	1685806	0,98	5,43	3,10E-04	2,50E-02
ENSGALG00000007945	CRYAB	24	6212347	6216461	-2,25	5,51	6,30E-05	6,60E-03
ENSGALG00000003212	TSPO2	26	4699001	4704104	-1,07	6,93	8,10E-06	1,10E-03
ENSGALG00000003485	CCND3	26	5031084	5043504	-0,78	5,45	6,00E-04	4,40E-02
ENSGALG00000033154	HOXB13	27	3743555	3745149	3,63	-0,43	4,00E-04	3,10E-02
ENSGALG00000003690	ENSGALG00000003690	27	5042180	5052819	0,9	6,98	6,30E-04	4,50E-02
ENSGALG00000042119	MIP	33	783896	785822	-3,46	2,29	1,80E-04	1,60E-02
ENSGALG00000034827	ENSGALG00000034827	AADN04000802.1	28868	41982	-8,32	1,16	6,70E-23	2,60E-20
ENSGALG00000042014	ENSGALG00000042014	AADN04000843.1	8281	15642	-8,46	2,52	1,80E-33	1,10E-30
ENSGALG00000035489	Wpkci-8	AADN04000843.1	16486	47731	-8,9	4,73	7,20E-90	1,00E-86
ENSGALG00000040531	ENSGALG00000040531	AADN04000843.1	25130	26410	-10,63	1,85	4,00E-28	2,00E-25
ENSGALG00000041221	ENSGALG00000041221	AADN04000843.1	36860	37655	-8,76	0,02	4,30E-08	9,30E-06
ENSGALG00000039438	ENSGALG00000039438	AADN04000879.1	34868	43489	-8,74	-0,02	3,30E-13	9,50E-11
ENSGALG00000043377	ENSGALG00000043377	AADN04001035.1	1532	3013	0,8	6,15	5,20E-04	3,90E-02
ENSGALG00000041555	ENSGALG00000041555	AADN04001035.1	4863	11120	0,73	8,01	1,10E-04	1,00E-02
ENSGALG00000029182	ENSGALG00000029182	AADN04001035.1	11904	18695	0,76	7,79	3,40E-05	4,10E-03
ENSGALG00000041887	ENSGALG00000041887	AADN04001119.1	29093	30246	-9,33	0,56	5,10E-18	1,70E-15
ENSGALG00000043000	ENSGALG00000043000	AADN04001516.1	10348	11870	-2,59	1,21	3,50E-04	2,80E-02
ENSGALG00000034597	faf	AADN04001826.1	2130	2901	-7,84	0,68	4,60E-19	1,60E-16
ENSGALG00000037690	ENSGALG00000037690	AADN04001826.1	3077	17579	-9,27	6,13	2,70E-14	8,00E-13
ENSGALG00000040470	ENSGALG00000040470	AADN04001826.1	17736	18882	-7,93	-0,8	1,60E-11	4,40E-09
ENSGALG00000038387	ENSGALG00000038387	AADN04001826.1	19951	21187	-6,23	-0,15	1,30E-11	3,60E-09
ENSGALG00000030729	ENSGALG00000030729	AADN04001957.1	14454	19983	1,01	4,52	5,90E-05	6,30E-03

ENSGALG00000040085	ENSGALG00000040085	AADN04002517.1	1941	5078	-7,65	0,56	1,30E-16	4,30E-14
ENSGALG00000043682	ENSGALG00000043682	AADN04002517.1	6076	7222	-6,29	-0,07	1,50E-13	4,50E-11
ENSGALG00000037738	ENSGALG00000037738	AADN04002626.1	480	14280	-8,12	2,54	4,20E-31	2,20E-28
ENSGALG00000035998	ENSGALG00000035998	AADN04002813.1	1981	16449	-10,86	6,21	1,00E-92	1,70E-89
ENSGALG00000039296	ENSGALG00000039296	AADN04003908.1	2274	5563	-7,08	4,88	4,20E-25	1,90E-22
ENSGALG00000043328	ENSGALG00000043328	AADN04003908.1	5719	6865	-9,58	2,41	5,30E-30	2,80E-27
ENSGALG00000029246	ENSGALG00000029246	AADN04003908.1	7857	10299	-8,68	3,46	5,30E-45	4,00E-42
ENSGALG00000041637	ENSGALG00000041637	AADN04004668.1	1046	4022	-9,15	0,33	3,30E-07	6,00E-05
ENSGALG00000032641	ENSGALG00000032641	AADN04004861.1	3875	12273	-8,28	-0,5	4,00E-10	9,90E-08
ENSGALG00000039978	SLC4A1	AADN04005854.1	5590	12646	-1,14	6,95	7,90E-05	7,80E-03
ENSGALG00000033782	ENSGALG00000033782	AADN04006411.1	731	6693	-9,61	2,45	8,30E-36	5,50E-33
ENSGALG00000036558	ENSGALG00000036558	AADN04006411.1	2208	10360	-8,9	0,12	7,00E-15	2,20E-12
ENSGALG00000036073	ENSGALG00000036073	AADN04006861.1	7170	9289	0,69	7,11	6,10E-04	4,50E-02
ENSGALG00000029890	ENSGALG00000029890	AADN04009627.1	199	3506	-8,45	4,02	7,40E-59	7,20E-56
ENSGALG00000033574	ENSGALG00000033574	AADN04017883.1	64	3185	-7,69	-1,01	4,00E-10	9,90E-08
ENSGALG00000035785	ENSGALG00000035785	AADN04024149.1	7018	9423	-10,3	3,88	2,70E-65	3,00E-62
ENSGALG00000036535	ENSGALG00000036535	AADN04024150.1	1	7941	-3,18	0,3	4,60E-08	9,70E-06
ENSGALG00000045359	ENSGALG00000045359	AADN04024150.1	4622	6901	-8,06	0,95	2,10E-21	7,90E-19
ENSGALG00000044781	ENSGALG00000044781	AADN04024151.1	2857	5136	-8,06	0,95	2,10E-21	7,90E-19
ENSGALG00000029783	ENSGALG00000029783	AADN04024629.1	4523	14980	-9,81	5,16	4,30E-117	8,90E-114
ENSGALG00000038856	ENSGALG00000038856	KQ759335.1	18952	50420	-7,21	1,7	3,10E-23	1,30E-20
ENSGALG00000030666	ENSGALG00000030666	KQ759365.1	30346	45877	-2,1	6,92	1,20E-06	2,00E-04
ENSGALG00000038891	ENSGALG00000038891	KQ759366.1	7614	8477	-9,28	0,51	1,50E-18	5,20E-16
ENSGALG00000035863	ENSGALG00000035863	KQ759486.1	65172	82484	-8,37	4,55	2,60E-45	2,10E-42
ENSGALG00000031015	ENSGALG00000031015	KQ759490.1	26121	39622	-8,04	0,89	1,60E-19	5,60E-17
ENSGALG00000032104	ENSGALG00000032104	KQ759493.1	3620	28747	-8,52	6,15	3,10E-156	2,20E-151
ENSGALG00000031712	ENSGALG00000031712	KQ759494.1	11522	14766	-8,74	4,07	2,60E-41	1,90E-38
ENSGALG00000040126	ENSGALG00000040126	KQ759494.1	15744	16890	-9,5	0,72	7,80E-20	2,80E-17
ENSGALG00000033705	ENSGALG00000033705	KQ759511.1	21	27908	-8,12	6,37	2,10E-151	9,90E-146
ENSGALG00000044800	ENSGALG00000044800	KQ759531.1	6339	38837	-2,83	2,11	2,40E-10	6,10E-08
ENSGALG00000033873	ENSGALG00000033873	KQ759531.1	27943	29013	-5,75	-0,62	6,70E-11	1,70E-08
ENSGALG00000034584	ENSGALG00000034584	Z	613361	742150	0,71	6,07	6,10E-05	6,50E-03
ENSGALG00000041052	ENSGALG00000041052	Z	852454	890244	0,72	6,57	1,90E-04	1,60E-02
ENSGALG00000031142	ACAA2	Z	985790	998134	0,65	6,97	5,00E-04	3,70E-02
ENSGALG00000036001	MADH2	Z	1515765	1555787	0,96	5,45	1,60E-06	2,50E-04
ENSGALG00000037880	ENSGALG00000037880	Z	2151465	2159381	1,02	9,49	2,10E-08	4,70E-06
ENSGALG00000035773	PIK3C3	Z	3992964	4061554	0,77	5,61	4,30E-05	4,90E-03
ENSGALG00000026547	TPGS2	Z	7103042	7126992	0,98	5,28	9,10E-06	1,20E-03
ENSGALG00000014168	ENSGALG00000014168	Z	7254494	7277712	1,08	4,88	5,80E-06	8,10E-04
ENSGALG00000030730	ENSGALG00000030730	Z	7277130	7287519	1,05	5,5	1,20E-06	2,00E-04
ENSGALG0000001668	ENSGALG0000001668	Z	7304115	7355704	0,68	6,55	1,70E-04	1,50E-02
ENSGALG00000013809	ENSGALG00000013809	Z	7358767	7475284	1,06	6,86	3,50E-09	8,50E-07
ENSGALG00000021365	DCTN3	Z	7914091	7921662	0,88	6,54	2,20E-06	3,40E-04
ENSGALG0000001986	ENSGALG0000001986	Z	8540992	8562186	0,97	8,85	1,10E-07	2,30E-05
ENSGALG00000027630	ENSGALG00000027630	Z	9095088	9106955	0,92	5,76	2,70E-06	4,10E-04
ENSGALG00000002394	RGP1	Z	9116332	9126132	0,91	4,85	5,50E-05	6,00E-03
ENSGALG00000002599	ARHGEF39	Z	9282639	9296314	0,96	4,48	4,30E-04	3,30E-02
ENSGALG00000002605	MRPL17	Z	9297783	9298741	0,82	5,68	1,10E-04	1,00E-02
ENSGALG000000028267	ENSGALG000000028267	Z	9668418	9696997	0,8	5,45	4,70E-05	5,30E-03
ENSGALG00000003288	TARS	Z	10102159	10116869	0,79	7,31	7,40E-06	1,00E-03
ENSGALG00000003373	BRIX1	Z	10563582	10567349	0,79	6,38	2,50E-05	3,00E-03

ENSGALG0000003605	ENSGALG0000003605	Z	11297074	11454131	1,02	7,42	1,90E-05	2,40E-03
ENSGALG0000003688	NUP155	Z	11533897	11564728	1,03	7,69	2,10E-07	4,00E-05
ENSGALG0000003708	WDR70	Z	11566340	11702225	1,02	5,94	5,00E-09	1,20E-06
ENSGALG0000004826	TTC33	Z	12854319	12891747	0,73	5,1	6,30E-04	4,50E-02
ENSGALG0000004874	ENSGALG0000004874	Z	14183746	14189286	1,01	4,52	5,90E-05	6,30E-03
ENSGALG00000034847	PELO	Z	15779060	15781904	0,84	5,47	1,50E-05	1,90E-03
ENSGALG00000020561	ENSGALG00000020561	Z	16555609	16594096	0,89	5,47	1,70E-04	1,50E-02
ENSGALG00000014709	SKIV2L2	Z	16760980	16805955	0,96	7	1,40E-07	2,70E-05
ENSGALG00000014740	IPO11	Z	19489356	19563300	0,67	6,5	1,70E-04	1,50E-02
ENSGALG00000014747	CWC27	Z	20211903	20325189	0,94	5,38	5,00E-06	7,10E-04
ENSGALG00000014753	CENPK	Z	20549127	20570556	0,75	5,68	6,30E-04	4,50E-02
ENSGALG00000014756	PPWD1	Z	20570641	20584993	0,72	6,36	7,20E-05	7,20E-03
ENSGALG0000004390	AP3B1	Z	22789995	22948814	0,83	6,24	5,10E-06	7,20E-04
ENSGALG00000023379	F2RL2	Z	23491744	23494829	1,07	3,47	4,40E-04	3,40E-02
ENSGALG00000028149	MED18	Z	23985560	23987165	0,87	4,97	1,10E-04	1,00E-02
ENSGALG00000014942	NSA2	Z	24202596	24206577	0,75	6,86	6,20E-05	6,50E-03
ENSGALG00000013510	UTP15	Z	24708064	24717704	0,72	6,15	6,60E-05	6,80E-03
ENSGALG00000005426	FREM1	Z	31717648	31775857	0,94	5,59	1,70E-04	1,50E-02
ENSGALG00000015091	HAUS6	Z	33583820	33602643	0,89	7,23	1,00E-06	1,70E-04
ENSGALG00000015090	PLIN2	Z	33603989	33617153	0,84	6,75	4,10E-06	5,90E-04
ENSGALG00000015082	RPS6	Z	33736457	33741021	0,72	10,78	2,40E-04	2,00E-02
ENSGALG00000015071	FOCAD	Z	34160761	34260236	0,96	6,59	4,10E-07	7,20E-05
ENSGALG00000015118	SMC5	Z	34954146	35009188	0,91	6	2,00E-06	3,20E-04
ENSGALG00000015144	ZFAND5	Z	35736063	35750582	0,83	6,77	4,80E-06	7,00E-04
ENSGALG00000015179	CEP78	Z	37936454	37958205	0,88	5,66	3,30E-06	4,90E-04
ENSGALG00000012591	HNRNPK	Z	40077218	40095705	0,74	9,08	4,20E-05	4,90E-03
ENSGALG00000012592	RMI1	Z	40095785	40099640	0,86	5,25	3,90E-05	4,60E-03
ENSGALG00000012598	NAA35	Z	40735008	40761607	0,84	6,05	2,30E-06	3,60E-04
ENSGALG00000012618	FANCC	Z	41661025	41748825	0,83	5,91	3,10E-04	2,50E-02
ENSGALG00000012621	ENSGALG00000012621	Z	41977884	42019651	0,78	5,32	1,30E-04	1,20E-02
ENSGALG00000014916	ENSGALG00000014916	Z	43221528	43247846	0,94	8,34	1,30E-07	2,60E-05
ENSGALG00000013493	SPTLC1	Z	44651919	44684718	0,82	5,81	1,20E-05	1,50E-03
ENSGALG00000029681	ENSGALG00000029681	Z	45700899	45716970	0,86	4,55	4,60E-04	3,50E-02
ENSGALG00000000218	ENSGALG00000000218	Z	46009581	46016260	0,72	6,2	1,20E-04	1,10E-02
ENSGALG00000000241	STARD4	Z	46509914	46518374	0,81	5,35	2,00E-04	1,70E-02
ENSGALG00000000247	WDR36	Z	46726620	46756993	0,98	6,62	1,50E-07	2,90E-05
ENSGALG00000015264	PPIP5K2	Z	49972661	50015273	0,79	5,86	3,70E-05	4,40E-03
ENSGALG00000015256	PAM	Z	50044582	50117715	0,69	5,96	1,20E-04	1,00E-02
ENSGALG00000014642	ENSGALG00000014642	Z	51274317	51349988	0,94	6,73	3,30E-06	4,90E-04
ENSGALG00000015290	LIX1	Z	51552825	51581108	1,91	4,05	6,60E-05	6,80E-03
ENSGALG00000015326	CLTA	Z	52040364	52055154	0,72	6,8	3,30E-04	2,60E-02
ENSGALG00000015349	PIGG	Z	53222477	53335831	0,94	5,02	4,70E-05	5,30E-03
ENSGALG00000015391	ENSGALG00000015391	Z	53776106	53829611	0,87	4,93	9,70E-05	9,40E-03
ENSGALG00000027334	ENSGALG00000027334	Z	54712941	54722162	0,8	5,8	9,70E-06	1,30E-03
ENSGALG00000014691	PHAX	Z	56181076	56190055	0,72	6,38	1,10E-04	1,00E-02
ENSGALG00000014670	TTC37	Z	57424796	57466657	0,8	6,33	9,60E-05	9,40E-03
ENSGALG00000027932	ENSGALG00000027932	Z	57809980	57821025	0,92	5,01	1,80E-05	2,30E-03
ENSGALG00000014661	FAM172A	Z	58092017	58355057	1,03	5,74	6,30E-07	1,10E-04
ENSGALG00000014648	CETN3	Z	59736248	59752830	0,87	5,11	3,10E-05	3,80E-03
ENSGALG00000015641	CCNH	Z	60998292	61007184	0,71	5,91	4,00E-04	3,10E-02
ENSGALG00000017706	ENSGALG00000017706	Z	61011125	61071967	0,93	6,25	1,50E-07	2,90E-05

ENSGALG00000027714	ATG10	Z	63371447	63454155	0,81	5,09	2,70E-04	2,20E-02
ENSGALG00000015603	ZCCHC9	Z	63736221	63742499	0,89	4,91	6,70E-05	6,90E-03
ENSGALG00000015589	MSH3	Z	63936017	64052808	1,03	5,35	3,60E-06	5,40E-04
ENSGALG00000020534	AK6	Z	64160615	64164655	1,09	6,02	2,30E-08	5,10E-06
ENSGALG00000015548	RNF20	Z	65079251	65097640	0,74	7,33	5,70E-05	6,20E-03
ENSGALG00000015663	HSDL2	Z	66024361	66036969	0,76	7	6,20E-05	6,50E-03
ENSGALG00000015691	SMC2	Z	66464325	66488192	1,03	8,9	4,00E-07	7,10E-05
ENSGALG00000001765	CAAP1	Z	68485992	68502940	0,95	4,43	1,90E-04	1,60E-02
ENSGALG00000001782	PLAA	Z	68502953	68521959	0,94	6,34	1,20E-07	2,40E-05
ENSGALG00000001801	IFT74	Z	68522460	68553748	0,96	4,66	4,30E-05	5,00E-03
ENSGALG00000001864	C9orf72	Z	68714996	68729199	0,83	4,61	5,50E-04	4,10E-02
ENSGALG00000001947	IKBKAP	Z	70551166	70582389	1,03	6,75	1,60E-08	3,60E-06
ENSGALG00000001954	APTX	Z	70584679	70593677	0,89	4,75	1,10E-04	1,00E-02
ENSGALG000000028998	XPA	Z	70731207	70737921	0,75	7,2	7,30E-05	7,30E-03
ENSGALG00000002187	HSD17B4	Z	71546997	71602452	0,91	6,73	2,50E-07	4,70E-05
ENSGALG00000008174	ENSGALG00000008174	Z	78589195	78621668	0,69	5,96	5,70E-04	4,20E-02
ENSGALG00000008229	ALDH7A1	Z	78926260	78944398	0,91	7,83	4,30E-07	7,50E-05
ENSGALG00000005316	SRFBP1	Z	80915551	80971412	1,04	5,3	3,50E-07	6,40E-05
ENSGALG00000005330	SNCAIP	Z	81090645	81151638	0,87	6,04	1,00E-05	1,40E-03
ENSGALG00000005346	PPIC	Z	81354026	81359622	0,87	5,41	3,70E-04	2,90E-02
ENSGALG000000029118	ENSGALG000000029118	Z	81574893	81580647	0,89	4,52	4,40E-04	3,40E-02
ENSGALG00000005395	POLR1E	Z	81615349	81627915	0,69	5,94	3,50E-04	2,80E-02
ENSGALG000000016438	MELK	Z	81980862	82003287	0,66	5,99	6,60E-04	4,70E-02

ID	logFC	nbDMC	DMC
ENSGALG00000001765	-11.08	1	DOWN
ENSGALG00000002187	-9.47	2	UP
ENSGALG00000003605	-9.03	2 DOWN	1 UP
ENSGALG00000003690	-9.29	1 DOWN	3 UP
ENSGALG00000003708	-10.06	1 DOWN	3 UP
ENSGALG00000003923	-9.15	1	DOWN
ENSGALG00000004390	-9.78	4	UP
ENSGALG00000005316	-10.52	2	UP
ENSGALG00000005330	-9.94	1	UP
ENSGALG00000005426	-10.28	2 DOWN	3 UP
ENSGALG00000006520	-8.56	1	UP
ENSGALG00000012592	-10.48	1	DOWN
ENSGALG00000012618	-10	2	UP
ENSGALG00000013809	-9.44	3	DOWN
ENSGALG00000014642	-9.48	4	UP
ENSGALG00000014661	-10.21	3 DOWN	3 UP
ENSGALG00000014916	-8.37	1	DOWN
ENSGALG00000015079	-10.56	1	UP
ENSGALG00000015391	-10.7	1	DOWN
ENSGALG00000017347	-4.39	3	UP
ENSGALG00000027170	-12.03	1	DOWN
ENSGALG00000027714	-10.57	2	DOWN
ENSGALG00000034584	-9.85	5 DOWN	3 UP
ENSGALG00000036001	-10.38	1	UP
ENSGALG00000036805	-8.94	5	DOWN
ENSGALG00000041052	-9.51	1 DOWN	3 UP
ENSGALG00000044240	-9.36	3	DOWN
ENSGALG00000044745	-9.54	1	UP

Table S7. Differentially methylated cytosines in gene body associated with differentially expressed genes between females and males (logFC > 0 : males more expressed).

Samples	Raw reads	Reads after mapping	Mapping rate (%)	Reads after cleaning	Duplication rate (%)
7127emb08	134142902	44009026	33.97	34685834	21.18
7127emb15	153903838	49659696	33.47	41457312	16.52
7127emb20	157230624	54914162	35.80	51168498	6.82
7127emb23	138736708	43542912	32.43	40951544	5.95
8303emb12	192625836	63143596	33.93	55582004	11.98
8303emb14	198175730	66115250	34.50	57012506	13.77
8303emb22	149396714	48339092	33.40	36114830	25.29

Table S8. Description of the global mapping analysis for each embryo against the diploid parental genome.

Samples	Detected CG	Depth	mCpG percent (%)
7127emb08	12764574	3.90	70.55
7127emb15	12655977	5.11	74.29
7127emb20	12533246	4.29	71.65
7127emb23	13207841	5.16	70.99
8303emb12	14879298	5.68	70.97
8303emb14	14628717	5.95	70.93
8303emb22	12370129	4.39	71.89

Table S9. Description of the cytosine position in CpG context and the global methylation analysis for each embryo in ASM.

Table S10. Allele-specific expression between alleles ($\log_{2}FC > 0$: line R⁻ more expressed). This long table is here : http://genoweb.toulouse.inra.fr/~mmersch/Article_Mersch-et-al/suppASE.pdf

References

1. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes & development* **23**, 781–3, DOI: [10.1101/gad.1787609](https://doi.org/10.1101/gad.1787609) (2009).
2. Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* **13**, 97–109, DOI: [10.1038/nrg3142](https://doi.org/10.1038/nrg3142) (2012).
3. Reik, W., Dean, W. & Walter, J. Epigenetic Reprogramming in Mammalian Development. *Epigenetics* **293**, 1089–1094 (2001).
4. Skinner, M. K. Environmental Epigenetics and a Unified Theory of the Molecular Aspects of Evolution: A Neo-Lamarckian Concept that Facilitates Neo-Darwinian Evolution. *Genome Biol. Evol.* **7**, 1296–1302, DOI: [10.1093/gbe/evv073](https://doi.org/10.1093/gbe/evv073) (2015).
5. Egger, G., Liang, G., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457–463, DOI: [10.1038/nature02625](https://doi.org/10.1038/nature02625) (2004).
6. Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. United States Am.* **105**, 17046–9, DOI: [10.1073/pnas.0806560105](https://doi.org/10.1073/pnas.0806560105) (2008).
7. Frésard, L. *et al.* Transcriptome-wide investigation of genomic imprinting in chicken. *Nucleic Acids Res.* **42**, 3768–3782, DOI: [10.1093/nar/gkt1390](https://doi.org/10.1093/nar/gkt1390) (2014).
8. Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40**, 904–908, DOI: [10.1038/ng.174](https://doi.org/10.1038/ng.174) (2008).
9. Tycko, B. Allele-specific DNA methylation: beyond imprinting. *Hum. Mol. Genet.* **19**, R210–R220, DOI: [10.1093/hmg/ddq376](https://doi.org/10.1093/hmg/ddq376) (2010).
10. Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20**, 883–889, DOI: [10.1101/gr.104695.109](https://doi.org/10.1101/gr.104695.109) (2010).
11. Cheung, W. A. *et al.* Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome biology* **18**, 50, DOI: [10.1186/s13059-017-1173-7](https://doi.org/10.1186/s13059-017-1173-7) (2017).
12. Bird, A. P. DNA methylation patterns and epigenetic memory. *Genes & Dev.* **16**, 6–21, DOI: [10.1101/gad.947102](https://doi.org/10.1101/gad.947102) (2002). [NIHMS150003](https://doi.org/10.1101/NIHMS150003).
13. Li, J. *et al.* Genome-wide DNA methylome variation in two genetically distinct chicken lines using MethylC-seq. *BMC Genomics* **16**, 851, DOI: [10.1186/s12864-015-2098-8](https://doi.org/10.1186/s12864-015-2098-8) (2015).
14. Zhang, Z. *et al.* Whole genome bisulfite sequencing reveals unique adaptations to high-altitude environments in Tibetan chickens. *PLOS ONE* **13**, e0193597, DOI: [10.1371/journal.pone.0193597](https://doi.org/10.1371/journal.pone.0193597) (2018).
15. Zhang, M. *et al.* Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Sci. Reports* **7**, 45564, DOI: [10.1038/srep45564](https://doi.org/10.1038/srep45564) (2017).
16. Vanyushin, B., Nemirovsky, L., Klimenko, V., Vasiliev, V. & Belozersky, A. The 5-Methylcytosine in DNA of Rats. *Gerontology* **19**, 138–152, DOI: [10.1159/000211967](https://doi.org/10.1159/000211967) (1973).
17. Hoal-van Helden, E. G. & van Helden, P. D. Age-related methylation changes in DNA may reflect the proliferative potential of organs. *Mutat. Res.* **219**, 263–266, DOI: [10.1016/0921-8734\(89\)90027-1](https://doi.org/10.1016/0921-8734(89)90027-1) (1989).
18. Bacalini, M. G. *et al.* The methylation of nuclear and mitochondrial DNA in ageing phenotypes and longevity. *Mech. Ageing Dev.* **165**, 156–161, DOI: [10.1016/j.mad.2017.01.006](https://doi.org/10.1016/j.mad.2017.01.006) (2017).
19. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492, DOI: [10.1038/nrg3230](https://doi.org/10.1038/nrg3230) (2012). [9605103](https://doi.org/10.1038/9605103).
20. Zhang, X. *et al.* Reduced representation bisulfite sequencing (RRBS) of dairy goat mammary glands reveals DNA methylation profiles of integrated genome-wide and critical milk-related genes. *Oncotarget* **8**, 115326–115344, DOI: [10.18632/oncotarget.23260](https://doi.org/10.18632/oncotarget.23260) (2017).
21. Deaton, A. M. & Bird, A. P. CpG islands and the regulation of transcription. *Genes & Dev.* **25**, 1010–1022, DOI: [10.1101/gad.2037511](https://doi.org/10.1101/gad.2037511) (2011).
22. Putiri, E. L. & Robertson, K. D. Epigenetic mechanisms and genome stability. *Clin. Epigenetics* **2**, 299–314, DOI: [10.1007/s13148-010-0017-z](https://doi.org/10.1007/s13148-010-0017-z) (2010).
23. Doerfler, W. DNA Methylation—A Regulatory Signal in Eukaryotic Gene Expression. *J. Gen. Virol.* **57**, 1–20, DOI: [10.1099/0022-1317-57-1-1](https://doi.org/10.1099/0022-1317-57-1-1) (1981).

24. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688, DOI: [10.1101/gr.136101.111](https://doi.org/10.1101/gr.136101.111) (2012).
25. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74, DOI: [10.1038/nature11247](https://doi.org/10.1038/nature11247) (2012). [1111.6189v1](https://doi.org/10.1101/6189v1).
26. Teif, V. B. *et al.* Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **24**, 1285–1295, DOI: [10.1101/gr.164418.113](https://doi.org/10.1101/gr.164418.113) (2014).
27. Schuyler, R. P. *et al.* Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems. *Cell Reports* **17**, 2101–2111, DOI: [10.1016/j.celrep.2016.10.054](https://doi.org/10.1016/j.celrep.2016.10.054) (2016).
28. Shao, X., Zhang, C., Sun, M.-A., Lu, X. & Xie, H. Deciphering the heterogeneity in DNA methylation patterns during stem cell differentiation and reprogramming. *BMC Genomics* **15**, 978, DOI: [10.1186/1471-2164-15-978](https://doi.org/10.1186/1471-2164-15-978) (2014).
29. Lee, S. T. *et al.* Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic Acids Res.* **43**, 2590–2602, DOI: [10.1093/nar/gkv103](https://doi.org/10.1093/nar/gkv103) (2015).
30. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10, DOI: [10.1186/gb-2011-12-1-r10](https://doi.org/10.1186/gb-2011-12-1-r10) (2011).
31. Teranishi, M. *et al.* Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the DMRT1 locus. *Chromosom. Res.* **9**, 147–165, DOI: [10.1023/A:1009235120741](https://doi.org/10.1023/A:1009235120741) (2001).
32. Bogani, D. *et al.* Loss of Mitogen-Activated Protein Kinase Kinase Kinase 4 (MAP3K4) Reveals a Requirement for MAPK Signalling in Mouse Sex Determination. *PLoS Biol.* **7**, e1000196, DOI: [10.1371/journal.pbio.1000196](https://doi.org/10.1371/journal.pbio.1000196) (2009).
33. Warr, N., Siggers, P., Carré, G.-A., Wells, S. & Greenfield, A. Genetic Analyses Reveal Functions for MAP2K3 and MAP2K6 in Mouse Testis Determination1. *Biol. Reproduction* **94**, 103–103, DOI: [10.1095/biolreprod.115.138057](https://doi.org/10.1095/biolreprod.115.138057) (2016).
34. Flores-Santin, J., Rojas Antich, M., Tazawa, H. & Burggren, W. W. Hematology from embryo to adult in the bobwhite quail (*Colinus virginianus*): Differential effects in the adult of clutch, sex and hypoxic incubation. *Comp. Biochem. Physiol. Part A: Mol. & Integr. Physiol.* **218**, 24–34, DOI: [10.1016/j.cbpa.2018.01.005](https://doi.org/10.1016/j.cbpa.2018.01.005) (2018).
35. Fair, J., Whitaker, S. & Pearson, B. Sources of variation in haematocrit in birds. *Ibis* **149**, 535–552, DOI: [10.1111/j.1474-919X.2007.00680.x](https://doi.org/10.1111/j.1474-919X.2007.00680.x) (2007).
36. Mank, J. E., Nam, K., Brunstrom, B. & Ellegren, H. Ontogenetic Complexity of Sexual Dimorphism and Sex-Specific Selection. *Mol. Biol. Evol.* **27**, 1570–1578, DOI: [10.1093/molbev/msq042](https://doi.org/10.1093/molbev/msq042) (2010).
37. Ayers, K. L. *et al.* RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* **14**, R26, DOI: [10.1186/gb-2013-14-3-r26](https://doi.org/10.1186/gb-2013-14-3-r26) (2013).
38. Izzo, A. *et al.* Dynamic changes in H1 subtype composition during epigenetic reprogramming. *The J. Cell Biol.* **216**, 3017–3028, DOI: [10.1083/jcb.201611012](https://doi.org/10.1083/jcb.201611012) (2017).
39. Wan, J. *et al.* Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics* **16**, 49, DOI: [10.1186/s12864-015-1271-4](https://doi.org/10.1186/s12864-015-1271-4) (2015).
40. Blattler, A. *et al.* Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* **15**, 469, DOI: [10.1186/s13059-014-0469-0](https://doi.org/10.1186/s13059-014-0469-0) (2014).
41. Frésard, L. *et al.* Genome-Wide Characterization of RNA Editing in Chicken Embryos Reveals Common Features among Vertebrates. *PLoS ONE* **10**, e0126776, DOI: [10.1371/journal.pone.0126776](https://doi.org/10.1371/journal.pone.0126776) (2015).
42. Snyder, E. M. *et al.* APOBEC1 complementation factor (AICF) is dispensable for C-to-U RNA editing in vivo. *RNA* **23**, 457–465, DOI: [10.1261/rna.058818.116](https://doi.org/10.1261/rna.058818.116) (2017).
43. Do, C. *et al.* Genetic–epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol.* **18**, 120, DOI: [10.1186/s13059-017-1250-y](https://doi.org/10.1186/s13059-017-1250-y) (2017).
44. Zhuo, Z., Lamont, S. J. & Abasht, B. RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken. *Sci. Reports* **7**, 11944, DOI: [10.1038/s41598-017-12179-9](https://doi.org/10.1038/s41598-017-12179-9) (2017).
45. Schröder, C. *et al.* Regions of common inter-individual DNA methylation differences in human monocytes: genetic basis and potential function. *Epigenetics & Chromatin* **10**, 37, DOI: [10.1186/s13072-017-0144-2](https://doi.org/10.1186/s13072-017-0144-2) (2017).

46. Palacios, R. *et al.* Allele-Specific Gene Expression Is Widespread Across the Genome and Biological Processes. *PLoS ONE* **4**, e4150, DOI: [10.1371/journal.pone.0004150](https://doi.org/10.1371/journal.pone.0004150) (2009).
47. Ellegren, H. *et al.* Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* **5**, 40, DOI: [10.1186/1741-7007-5-40](https://doi.org/10.1186/1741-7007-5-40) (2007).
48. Itoh, Y. *et al.* Sex bias and dosage compensation in the zebra finch versus chicken genomes: General and specialized patterns among birds. *Genome Res.* **20**, 512–518, DOI: [10.1101/gr.102343.109](https://doi.org/10.1101/gr.102343.109) (2010).
49. Bumstead, N. & Barrow, P. A. Genetics of Resistance to Salmonella Typhimurium in Newly Hatched Chicks. *Br. Poult. Sci.* **29**, 521–529, DOI: [10.1080/00071668808417078](https://doi.org/10.1080/00071668808417078) (1988).
50. Bordas, A., Tixier-Boichard, M. & Merat, P. Direct And Correlated Responses To Divergent Selection For Residual Food Intake In Rhode Island Red Laying Hens. *Br. Poult. Sci.* **33**, 741–754, DOI: [10.1080/00071669208417515](https://doi.org/10.1080/00071669208417515) (1992).
51. Warren, W. C. *et al.* A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda, Md.)* **7**, 109–117, DOI: [10.1534/g3.116.035923](https://doi.org/10.1534/g3.116.035923) (2017).
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, DOI: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) (2009). [1006.1266v2](https://doi.org/10.1093/bioinformatics/btp352).
53. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303, DOI: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110) (2010). [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
54. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92, DOI: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695) (2012).
55. Andrews, S. R. FastQC: a quality control tool for high throughput sequence data (2010).
56. Djebali, S. *et al.* Bioinformatics Pipeline for Transcriptome Sequencing Analysis. vol. 1468 of *Methods in Molecular Biology*, 201–219, DOI: [10.1007/978-1-4939-4035-6_14](https://doi.org/10.1007/978-1-4939-4035-6_14) (Springer New York, New York, NY, 2017).
57. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, DOI: [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) (2013). [1201.0052](https://doi.org/10.1093/bioinformatics/bts635).
58. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329, DOI: [10.1093/bioinformatics/btr355](https://doi.org/10.1093/bioinformatics/btr355) (2011). [171](https://doi.org/10.1093/bioinformatics/btr355).
59. Robinson, M. D. *et al.* Evaluation of affinity-based genome-wide DNA methylation data: Effects of CpG density, amplification bias, and copy number variation. *Genome Res.* **20**, 1719–1729, DOI: [10.1101/gr.110601.110](https://doi.org/10.1101/gr.110601.110) (2010).
60. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566, DOI: [10.1038/nprot.2013.092](https://doi.org/10.1038/nprot.2013.092) (2013).
61. Krueger, F., Kreck, B., Franke, A. & Andrews, S. R. DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods* **9**, 145–151, DOI: [10.1038/nmeth.1828](https://doi.org/10.1038/nmeth.1828) (2012).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25, DOI: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25) (2009).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, DOI: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) (2009). [1303.3997](https://doi.org/10.1093/bioinformatics/btp324).
64. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87, DOI: [10.1186/gb-2012-13-10-r87](https://doi.org/10.1186/gb-2012-13-10-r87) (2012).
65. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42**, 1–11, DOI: [10.1093/nar/gku154](https://doi.org/10.1093/nar/gku154) (2014).
66. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–22, DOI: [10.1038/nature08514](https://doi.org/10.1038/nature08514) (2009).
67. Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res.* **43**, 1–9, DOI: [10.1093/nar/gkv715](https://doi.org/10.1093/nar/gkv715) (2015).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158, DOI: [10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330) (2011). [NIHMS150003](https://doi.org/10.1093/bioinformatics/btr330).



FIGURE 3.1 – Visualisation des lectures alignées sur le chromosome W. Grâce au navigateur de génome IGV, nous pouvons visualiser des lectures alignées sur une région du chromosome W pour les sept embryons. On voit que l’embryon 8303emb18 a un profil de lectures proche des embryons femelles alors qu’il s’agit d’un mâle.

Embryons	Sexe	chr W	% lectures chrW	chr Z	% lectures chrZ
7127emb08	femelle	54014	1,87%	2891478	4,37%
7127emb15	mâle	5068	0,08%	6527204	8,52%
7127emb20	mâle	5442	0,06%	9753900	8,36%
7127emb23	femelle	65042	2,05%	3166532	4,37%
8303emb12	femelle	90432	1,95%	4625724	4,45%
8303emb14	mâle	8044	0,09%	9044002	8,30%
8303emb18	mâle	75976	1,68%	4528730	4,42%
8303emb22	femelle	52548	1,72%	3057560	4,55%

Tableau 3.1 – Nombre de lecture alignées par embryon sur les chromosomes sexuels. Les femelles ont un nombre qui varie entre 52548 à 90432 sur le chromosome W tandis que les lectures des embryons mâles sur le chromosome W varient entre 5068 et 8044. L’embryon 8303emb18, mâle, a un nombre de lecture proche des femelles.

3.2 Analyses complémentaires

3.2.1 Évolution des données

3.2.1.1 Évolution des versions du génome

Au cours de la thèse, la version du génome de référence utilisé pour les analyses a changé, passant de la version 4 à la version 5 (cf Tableau 1.2 page 60). Dans la dernière version, la taille de contigs augmente et quand ils s'assemblent, leur nombre diminue. Cependant, le nombre de *scaffolds* augmente et leur taille diminue. En effet, les *scaffolds* récupèrent des séquences non placées de la version précédente qui étaient riches en séquences répétées et donc plus difficiles à assembler. Les *scaffolds* nouvellement obtenus sont donc plus petits et correspondent à l'assemblage de très petits contigs de la version précédente [153]. En conséquence, les analyses des données WGBS ont été refaites sur la nouvelle version, et chaque analyse intègre les spécificités de la version utilisée (noms des chromosomes et *scaffolds* notamment). Le taux d'alignement des lectures sur les deux versions des génomes est en moyenne de 67,49% et 70,37% respectivement pour la version 4 (galgal4) et la version 5 (galgal5) du génome de *Gallus gallus*. Cela représente plus de lectures alignées sur la dernière version, ce qui, en conséquence, impacte le nombre de CG détectés. Le génome compte 10716562 CG sur la version 4 et 17888534 CG sur la version 5 (nombre d'occurrences de CG que nous avons calculé sur les génomes de référence). En moyenne, sur nos données WGBS, 10,29 millions de CG sont détectés sur la version 4 (96,05%) et 12,46 millions de CG sur la version 5 (69,65%), avec une profondeur de 1X. De plus, le taux de méthylation sur ces CG est légèrement moins haut sur galgal4 (63,60%) que sur galgal5 (66,74%), avec une profondeur moyenne de 8,20X (galgal4) et 9,22X (galgal5). En outre, le taux de conversion des C hors CG (cf partie 3.2.2.2 page 111) est supérieur à 99% dans les deux versions (99,28% pour galgal4 et 99,23% pour galgal5), ce qui est attendu : le génome de référence utilisé ne doit pas impacter le taux de conversion estimé. Ceci confirme l'efficacité de la conversion au bisulfite. Ainsi, la dernière version du génome de référence, mieux assemblée, permet de détecter plus de CG avec une profondeur légèrement plus grande. La version 6 de *Gallus gallus* est en cours d'annotation pour être mise en ligne par le navigateur *Ensembl*. Quand elle sera disponible, il sera possible de ré-utiliser le même pipeline pour analyser le méthylome en fonction des caractéristiques géniques.

3.2.1.2 Évolution du jeu de données

Le nombre d'échantillons à analyser est passé de 8 à 7 au cours de ma thèse. En effet, parmi les 4 femelles et les 4 mâles du dispositif WGBS, un mélange d'ADN entre une femelle et un mâle a été détecté. Un des individus identifié mâle a été contaminé par de l'ADN femelle. En effet, après avoir aligné les lectures, un nombre anormal de celles-ci étaient présentes sur le chromosome W, spécifique de la femelle (Figure 3.1). Quand on observe le nombre de lectures qui s'alignent sur les chromosomes W et Z, l'embryon 8303emb18, noté "mâle", a un nombre de lectures similaire aux femelles (Tableau 3.1). Afin de conclure sur le sexe de cet embryon, le sexage de l'individu à partir du tube d'ADN d'origine a été refait et confirme que c'est un mâle. Nous soupçonnons un mélange d'ADN lors de la préparation de la librairie WGBS. De plus, les analyses de PCA ont montré une forte ressemblance avec 8303emb12 qui est une femelle, ce qui laisse supposer que l'ADN femelle de 8303emb12 a été mélangé avec celui du mâle 8303emb18. La réciproque n'est pas vraie : nous avons vérifié sur plusieurs sites polymorphes que les données issues de l'embryon 8303emb12 correspondait bien à l'individu 8303emb12.

Cet embryon 8303emb18 a été supprimé lors du changement dans l'utilisation du génome de référence (de galgal4 à galgal5). La description des données précédentes sur la version 4 de *Gallus gallus* est faite sur huit embryons et les résultats présentés par la suite sur galgal5, et dans le projet de publication (cf partie 3.1 page 83), sont analysés à partir de sept embryons.

3.2.2 Contrôle de la qualité des données

3.2.2.1 Des lectures de séquençage aux CG "brutes"

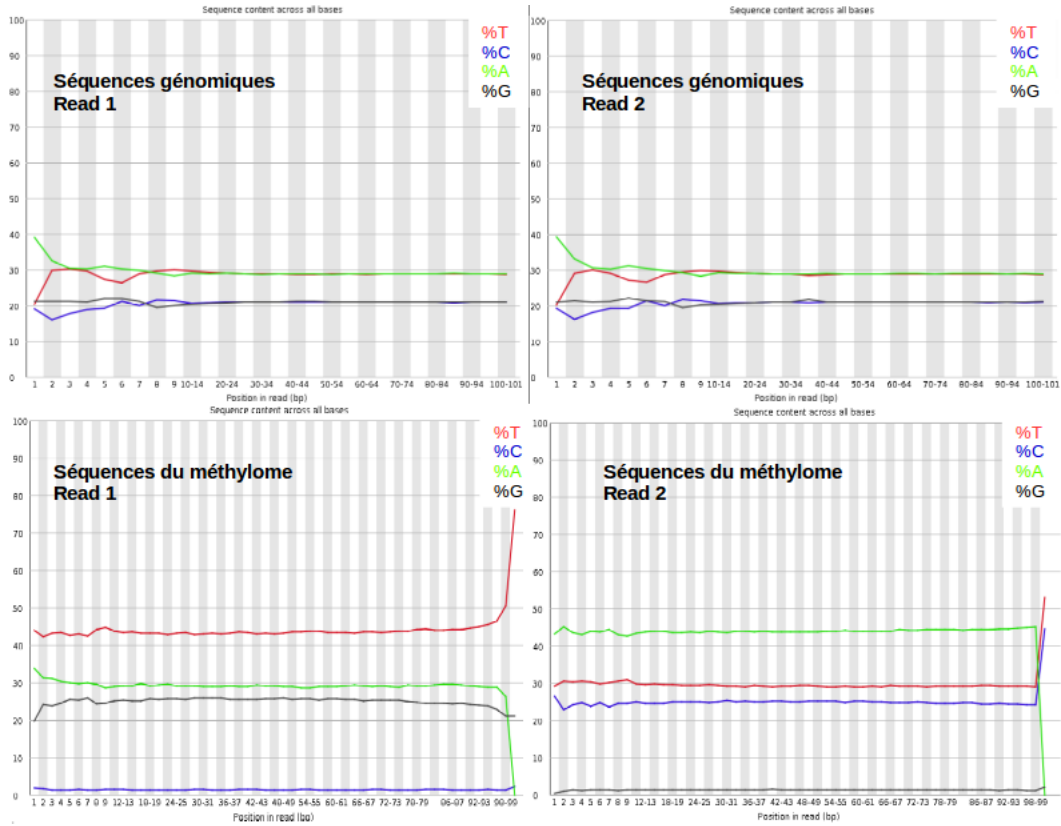


FIGURE 3.2 – Composition en bases sur l'ADN, d'abord sur les lectures génomiques non traitées au bisulfite de sodium puis sur celles traitées au bisulfite de sodium. Les brins sens (*read 1*) et anti-sens (*read 2*) de l'ADN sont représentés.

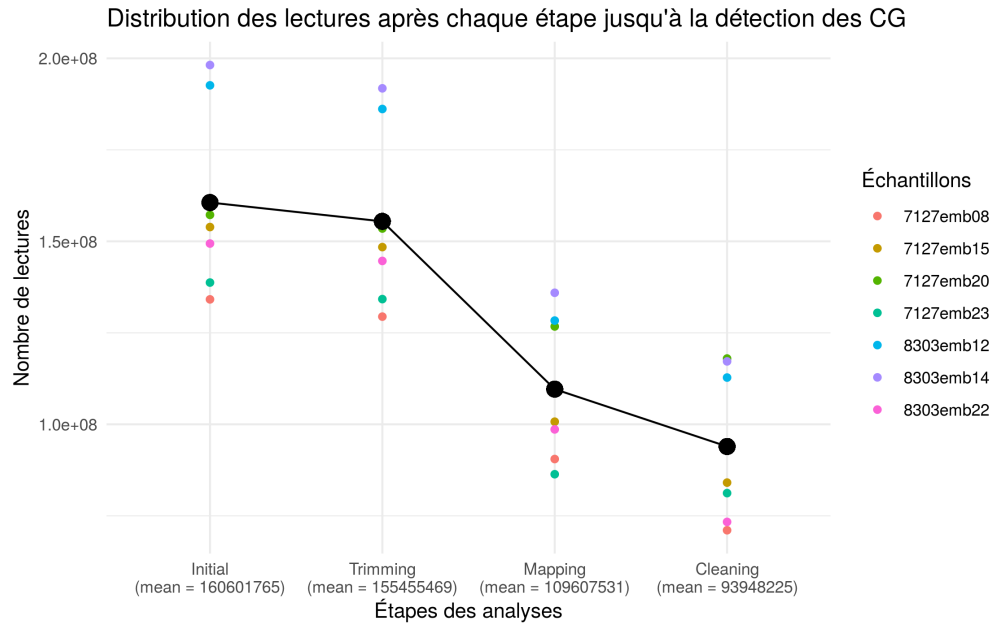


FIGURE 3.3 – Distribution des lectures à chaque étape du pipeline. L'axe x représente les étapes dans l'ordre du pipeline et l'axe y est le nombre de lectures par embryon. Les points noirs représentent la moyenne des valeurs des sept embryons.

Embryons	Initial	Après nettoyage	Après alignement	Après filtres
7127emb08	100%	96,48%	67,48%	52,97%
7127emb15	100%	96,44%	65,44%	54,63%
7127emb20	100%	97,63%	80,61%	75,04%
7127emb23	100%	96,74%	62,25%	58,53%
8303emb12	100%	96,64%	66,63%	58,55%
8303emb14	100%	96,79%	68,60%	59,14%
8303emb22	100%	96,82%	66,01%	49,09%
Moyenne	100%	96,79%	68,15%	58,28%

Tableau 3.2 – Pourcentage de séquences et de lectures restantes à travers les étapes successives d’analyses bioinformatiques.

À l’issue du nettoyage des séquences d’adaptateurs, l’outil FastQC permet d’évaluer les profils des séquences nettoyées. La composition en bases à chaque position des lectures, pour un embryon, sur les deux brins traités au bisulfite de sodium (méthylome) est différente de celle sur les brins d’ADN correspondant (ADN génomique non traité) (Figure 3.2). Sur les séquences génomiques, il y a environ 30% de A, 20% de C, 20% de G et 30% de T. Dans le génome galgal5, il y a 28,32% de A, 20,97% de C, 21,47% de G, 28,28% de T et 0,96% de bases non identifiées. Les observations sur les graphiques des séquences génomiques de FastQC correspondent bien au taux de CG attendu. Lorsque le traitement chimique au bisulfite de sodium a bien fonctionné, le pourcentage de C doit diminuer sur le brin sens au profit du pourcentage de T (*read 1*), et le pourcentage de A pour les lectures correspondant au brin anti-sens (*read 2*) doit augmenter aux dépens du pourcentage de G. Sur nos données WGBS, après conversion au bisulfite, le nombre de C tombe à 1% et le nombre de A augmente à plus de 40%. Cela confirme que la majorité des C du génome sont non méthylés, et signifie que le traitement au bisulfite a correctement remplacé les C non méthylés par des T.

Sur l’ensemble des analyses bioinformatiques à partir des données initiales jusqu’à la détection des CG, environ 41,72% de lectures ont été filtrées (Figure 3.3). On observe que l’embryon 7127emb20 s’aligne mieux que les six autres embryons (Tableau 3.2). Cela est probablement dû aux deux préparations de librairies qui n’ont pas eu lieu en même temps. Dans un premier temps, l’embryon 7127emb20 a servi d’embryon test à la mise en place du protocole WGBS puis plus tard, les six autres embryons ont été regroupés en pool lors de la préparation de la librairie. Il est probable que la première préparation était de meilleure qualité (meilleure intégrité de l’ADN, moins de lectures correspondant à un contaminant comme des dimères d’amorces,...). Les résultats sur la détection des CG sont similaires dans les sept embryons (cf projet de publication, partie 3.1 page 83).

3.2.2.2 Contrôle de la conversion au bisulfite

Un premier point de contrôle de la conversion au bisulfite est de regarder la conversion des C non méthylés en T. Si plus de 99% des C sont convertis en T alors la conversion a bien fonctionné. Le taux de conversion des C hors CG est calculé par le nombre de T (c’est-à-dire C non méthylé) divisé par la couverture pour chaque C hors contexte CpG. Le taux moyen de conversion donné est de 99,23% sur nos données WGBS (Tableau 3.3). Cependant, le taux de conversion des six embryons n’est pas de 100%. Cette faible différence est due au fait que moins de 1% des C non méthylés du génome n’ont pas été convertis (sous-conversion), ce qui est observé dans toutes les expériences d’analyses de méthylation par séquençage après bisulfite de sodium. Un deuxième point de contrôle du traitement au bisulfite est la conversion inappropriée des C méthylés en T. Pour détecter cette sur-conversion, et confirmer la sous-conversion, un plasmide circulaire de 2686 pb a été ajouté pendant la préparation des librairies, avant le traitement au bisulfite de sodium. Les CG du plasmide sont totalement méthylés. Pour détecter dans les fragments d’ADN du plasmide, le génome du plasmide a été ajouté au génome de référence lors des analyses de la partie 1 du pipeline (partie "Alignement et filtres"). L’alignement contre le génome de référence concaténé avec le génome du plasmide a permis de détecter les CG issus du plasmide. Lors de l’analyse du taux de méthylation sur le génome du plasmide de ces CG dans les contextes CpG, CHG

Embryons	Taux de conversion
7127emb08	99,23%
7127emb15	99,25%
7127emb20	99,49%
7127emb23	98,96%
8303emb12	99,22%
8303emb14	99,23%
8303emb22	99,25%
Moyenne	99,23%

Tableau 3.3 – Taux de conversion des C hors CG pour tous les embryons.

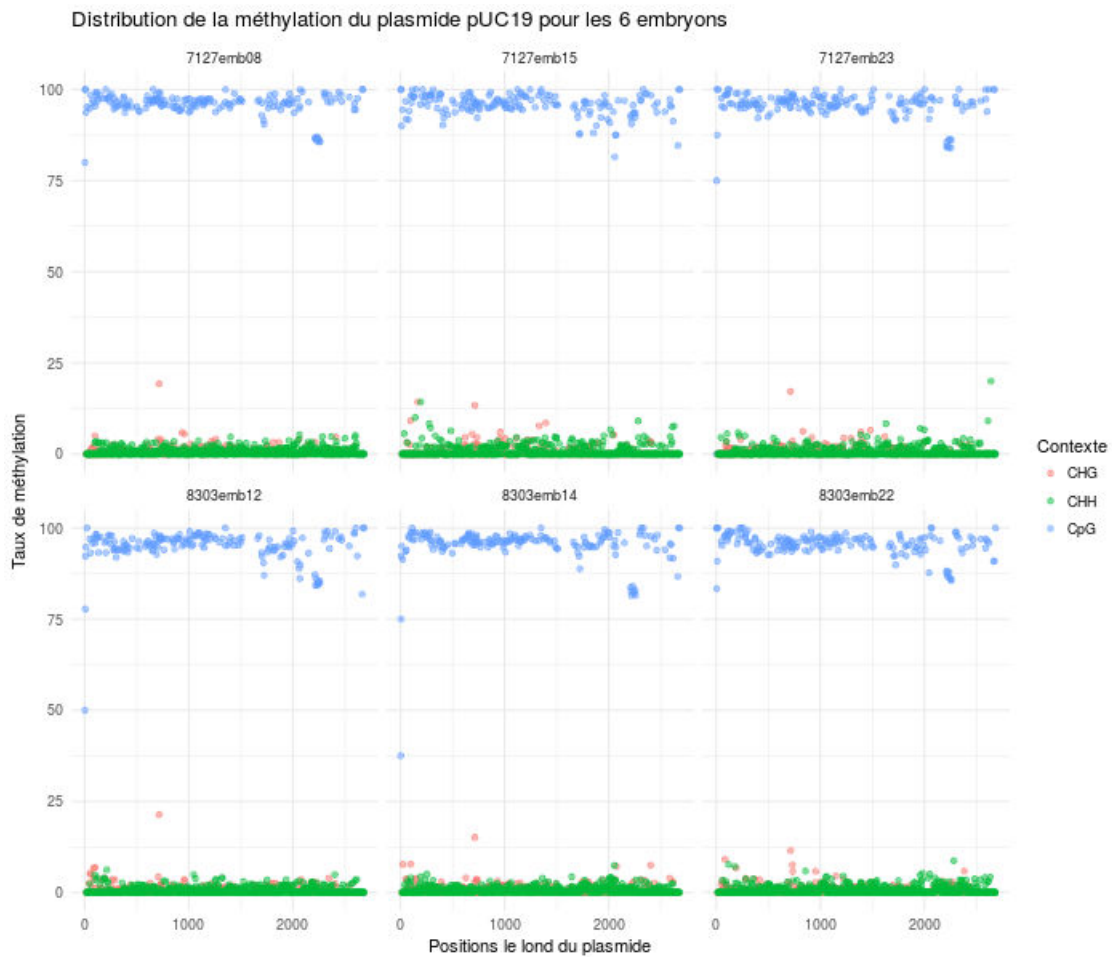


FIGURE 3.4 – Taux de méthylation des C du plasmide en WGBS. Chaque couleur représente un contexte de C : CpG, CHG et CHH. Toutes les cytosines en contexte CpG sont méthylées, autour de 100%, tandis que les cytosines dans les deux autres contextes sont très peu méthylées.

et CHH, un taux de méthylation est calculé (Figure 3.4). Sur les six embryons, les cytosines en contexte CpG sont très méthylées (94,12%) et les cytosines hors CG sont peu méthylées (0,78% pour CHG et 0,64% pour CHH) (Tableau 3.4). Le profil de conversion de l’embryon 7127emb20 montre un mélange de plasmide méthylé et non méthylé, et donc n’est pas informatif pour l’analyse du plasmide. Il n’est donc pas représenté dans les analyses du plasmide. Les deux contextes hors CG ont plus de 99% de conversion des C non méthylés, ce qui est cohérent avec le premier point de contrôle. Dans le contexte CpG, la différence entre 100% et le taux de conversion observé est la sur-conversion : des C qui devraient rester C sont convertis par le bisulfite. Cette sur-conversion induit une légère sous-estimation du taux de méthylation [172].

Embryons	Contexte CpG	Contexte CHG	Contexte CHH
7127emb08	93,88%	0,78%	0,63%
7127emb15	95,29%	0,74%	0,69%
7127emb23	93,61%	0,68%	0,56%
8303emb12	93,35%	0,79%	0,55%
8303emb14	94,18%	0,92%	0,69%
8303emb22	94,45%	0,79%	0,70%
Moyenne	94,12%	0,78%	0,64%

Tableau 3.4 – Taux de méthylation moyen du plasmide. La distribution du taux de méthylation moyen des C est représentée pour les trois contextes CpG, CHG et CHH du plasmide par embryon.

3.2.3 Choix des paramètres d’analyses

3.2.3.1 Définition d’un minimum de couverture

Définir un minimum de couverture sur les données permet de s’assurer de la pertinence des taux de méthylation observés. Par exemple si un CG a un couverture de 2X avec une lecture C et l’autre T, le taux de méthylation est de 50% dans cet embryon. Si sur cette même position, dans un autre embryon, la couverture est également de 2X mais 2 C sont présents et 0 T, le taux de méthylation est de 100%. Les résultats sur la détection des taux de méthylation différentielle pourraient ainsi détecter cette position comme étant différentielle (50% *versus* 100%), à la significativité près, alors que la faible couverture ne permet pas d’en être sûre. Sur nos données, les couvertures des CG sans filtres ne sont pas complètement homogènes entre embryon (Figure 3.5). La médiane de couverture la plus faible est de 5X pour 7127emb08, 7127emb15 et 8303emb22. Pour ce type d’analyses WGBS, une profondeur minimale de 5X est recommandée, avec un nombre de réplicats biologiques au moins égal à 2 [192]. Nous avons choisi donc d’appliquer sur les positions communes aux 7 embryons un minimum de couverture de 5X dans au moins 75% des échantillons, soit 5/7.

3.2.3.2 Méthodes d’annotation

Attribuer à chaque CG détecté une position dans une région génomique fonctionnelle permet de caractériser les niveaux de méthylation le long du génome. Deux méthodes sont proposées : soit la position est vue une seule fois (présence dans une seule catégorie fonctionnelle), c’est la méthode par projection, soit la position est vue plusieurs fois (peut être dans plusieurs catégories fonctionnelles), c’est la méthode par redondance.

La méthode par projection nécessite de créer une priorité entre les régions fonctionnelles. Dans l’annotation 89 de *Ensembl* sur le génome de référence *Gallus gallus* version 5, huit catégories fonctionnelles sont définies : gène, transcrit, CDS (*Coding-Sequence*), exon, 5’UTR (*UnTranslated Region* en 5’), 3’UTR (*UnTranslated Region* en 3’), codon start (démarrage de la traduction) et codon stop (arrêt de la traduction). L’annotation définit un intervalle fonctionnel sur le génome, le gène. La région en amont du gène est connue pour être indispensable à la régulation de la transcription, il s’agit du promoteur. Nous définissons le promoteur comme la région de 1 kb en amont du TSS (*Transcription Start Site*), valeur souvent retrouvée dans la littérature et correspondant à une région chez le poulet où le niveau de méthylation se stabilise par rapport au TSS

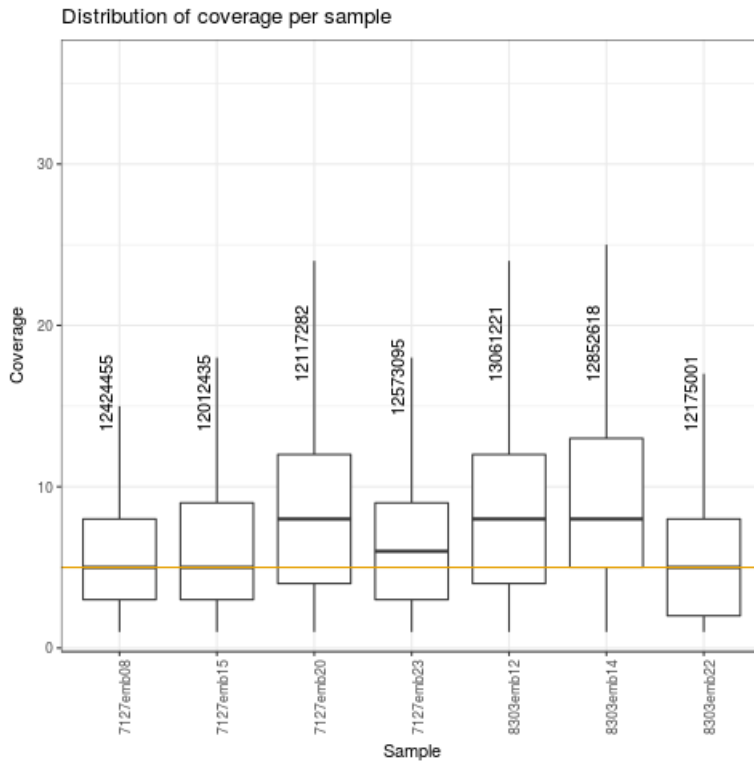


FIGURE 3.5 – Distribution de la couverture des CG dans chaque embryon. La ligne orange représente le minimum de couverture appliqué, soit une couverture de 5X.

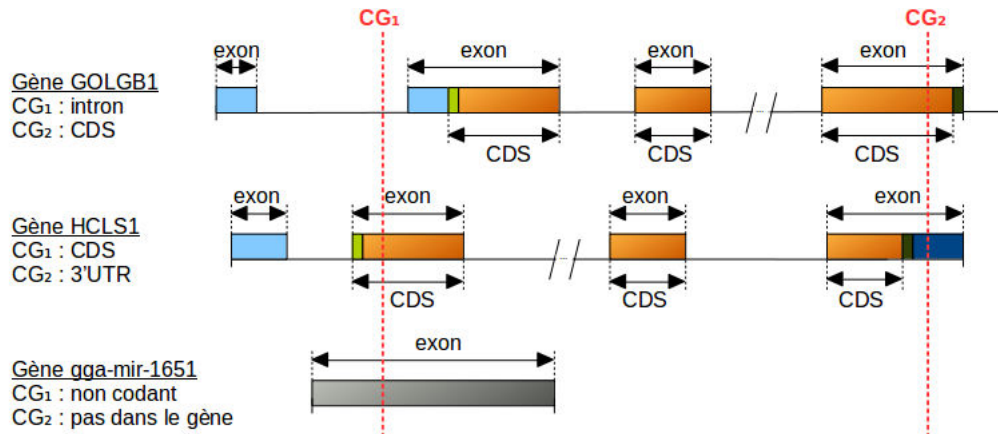


FIGURE 3.6 – Schéma de deux CG dans les caractéristiques géniques. La couleur bleu clair représente la région 5'UTR, la couleur vert clair le codon start, la couleur orange le CDS, la couleur vert foncé le codon stop, la couleur bleu foncé la région 3'UTR et la couleur grise le non codant. Lorsqu'on projette sur le génome, le CG₁ est présent dans l'intron du gène GOLGB1, le CDS du gène HCLS1 et dans le non codant dans le gène gga-mir-1651. Selon l'ordre de priorité défini, ce CG est dans la région CDS (CDS > intron > non codant). De même, le CG₂ est présent dans le codon stop dans le gène GOLGB1, dans la région 3'UTR dans le gène HCLS1 et n'est pas dans le gène gga-mir-1651. Le codon stop étant regroupé avec la région 3'UTR, le CG projeté est présent dans le 3'UTR.

[193]. La priorité de la position du CG est donc d'abord le gène (car annoté) puis le promoteur. Si le CG n'est dans aucune de ces deux régions alors il est considéré comme intergénique. De plus, un gène est défini comme l'ensemble des exons et introns (régions codantes et non codantes) et des régions 5' et 3' non traduites (5'UTR et 3'UTR). L'ordre de priorité est défini comme suit : le CDS associé avec le codon start, puis la région 5'UTR et la région 3'UTR associée avec le codon stop. L'annotation exon n'est pas utilisée pour définir les exons "projetés" car elle peut regrouper des parties codantes et non codantes. Cependant, elle est utilisée pour déduire la région intronique. Enfin, tout ce qui n'est pas annoté précédemment par projection est noté région non codante. Ainsi, deux groupes se dessinent (Figure 3.6) : un groupe de caractéristiques génomiques, avec les gènes, promoteurs et les régions intergéniques et un groupe de caractéristiques géniques (au sein du gène) que sont les annotations CDS, 5'UTR, 3'UTR, intron et non codant.

Une autre catégorie fonctionnelle peut être définie : les îlots CpG, ou CGI. Dans l'annotation générée par le logiciel EMBOSS (version 6.4.0.0), un CGI est défini comme ayant un taux de CG de plus de 50%, un ratio de CpG observé / CpG attendu supérieur à 0,6 et une taille minimum de 200 pb. 59116 CGI sont projetés sur le génome et sont la première priorité. Au sein de cette catégorie fonctionnelle, les régions flanquantes des CGI sont également représentées. Les régions flanquantes de 1 kb autour du CGI sont appelées *shore* et les régions flanquantes de 2 kb sont notées *shelf* (Figure 3.7). Le reste est considéré comme de l'*openSea*.

La deuxième méthode pour annoter les CG peut être d'attribuer plusieurs régions fonctionnelles au CG. Par exemple, dans un gène, le CG peut être dans un exon et dans un autre gène, le CG peut être dans le 5'UTR (Figure 3.8). Cette méthode permet de détecter un plus grand nombre de CG dans chaque région fonctionnelle. Sur galgal5, 21391 gènes projetés sont dénombrés (fusion des intervalles de gènes) contre 24881 des gènes présents dans l'annotation d'*Ensembl* (positions des gènes indépendantes les unes des autres). De même, dans les promoteurs, nous avons trouvé 20439 promoteurs "projetés" et 24872 promoteurs issus de la méthode redondance. La différence entre les 24872 promoteurs "redondants" obtenus et les 24881 promoteurs attendus vient du fait que 3 *scaffolds* commencent au TSS de leur gène (deux à la position du TSS et un à la position 1 du TSS) et 6 gènes différents ont le même TSS donc leurs 6 promoteurs sont assimilés à un seul promoteur. Cependant, la répartition des CG dans l'annotation est similaire entre les deux méthodes (Figure 3.9).

3.2.3.3 Méthylation au TSS

Comme montré dans l'article, le niveau de méthylation diminue avant le TSS. À une échelle différente, on voit que la répartition de la méthylation entre -1000 et +1000 pb autour du TSS a un pic proche du TSS (Figure 3.10a). Ce pic n'est pas retrouvé dans la littérature et nous nous sommes interrogés sur sa nature : réalité biologique ou artefact ? En raison de la faible profondeur de séquençage, peu de CG sont détectés autour du TSS. Trois groupes de gènes ont été constitués en fonction de la densité des CG autour du TSS : très denses ("fort") en CG, peu denses ("faible"), ou sans CG ("absent") autour du TSS (Figure 3.10 b). On observe que les gènes riches en CG sont également ceux pour lesquels le taux de méthylation est le plus fort autour du TSS. Ce sont également ceux à partir desquels est calculé le niveau moyen de méthylation dans l'intervalle -300 pb / +300 pb autour du TSS, où est le pic. Ces résultats semblent signifier que le pic observé pour l'ensemble des gènes est un pic artefactuel lié à la faible profondeur de séquençage de notre jeu de données. En effet, l'hypothèse la plus probable est que les gènes dont on ne détecte pas les CG dans l'intervalle -300 pb / +300 pb ont un taux de méthylation plus faible que les autres gènes, en accord avec les taux de méthylation observés de part et d'autre de cet intervalle. La moyenne de méthylation pourrait donc être sur-évaluée.

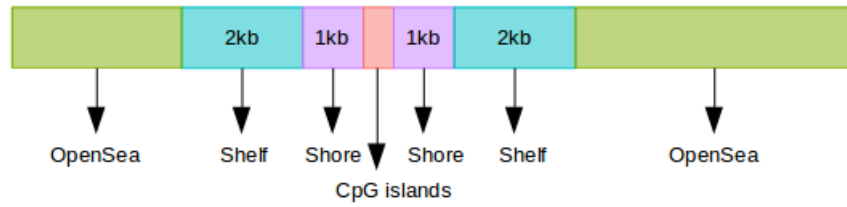


FIGURE 3.7 – Schéma d'un îlot CpG, des *shores*, des *shelves* et de l'*openSea*.

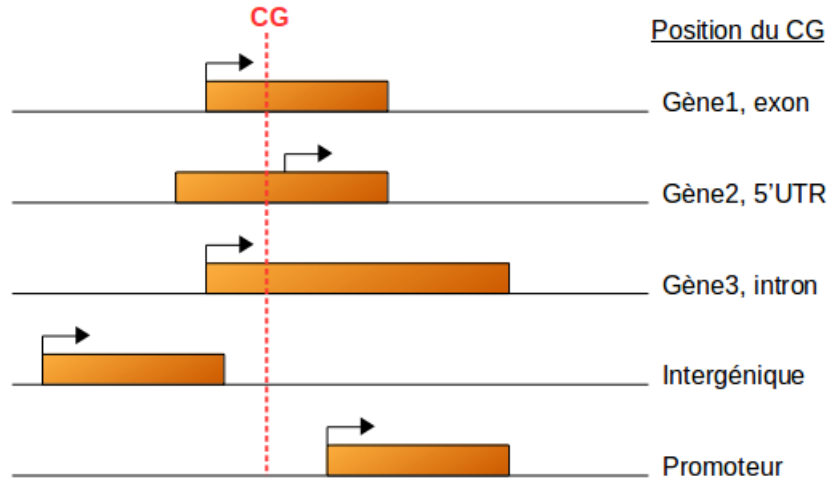
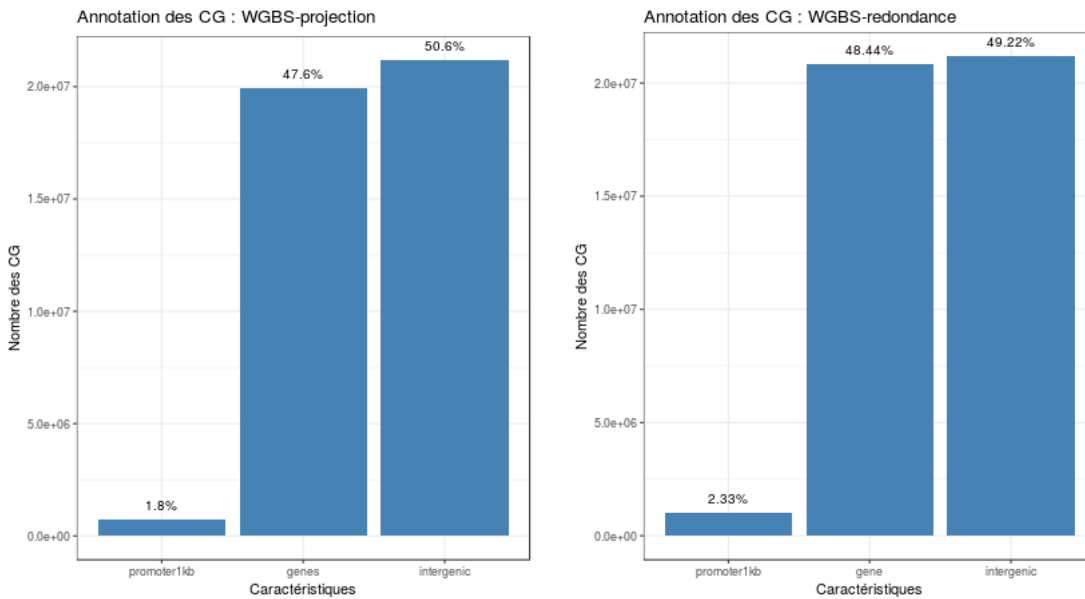


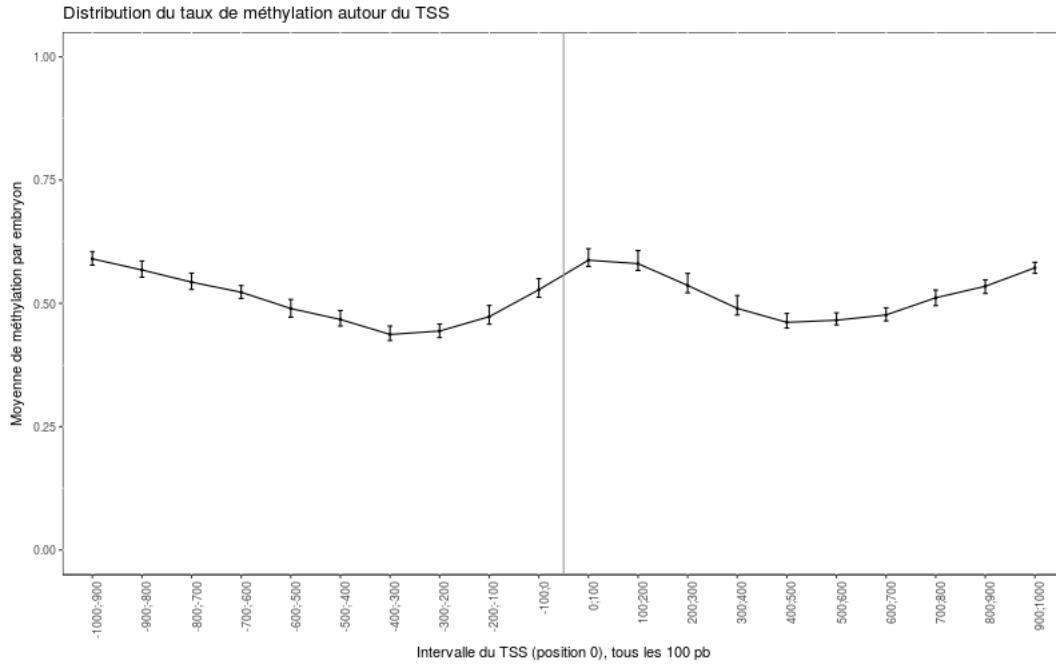
FIGURE 3.8 – Schéma d'une même position de CG dans plusieurs régions génomiques. La couleur orange représente le gène.



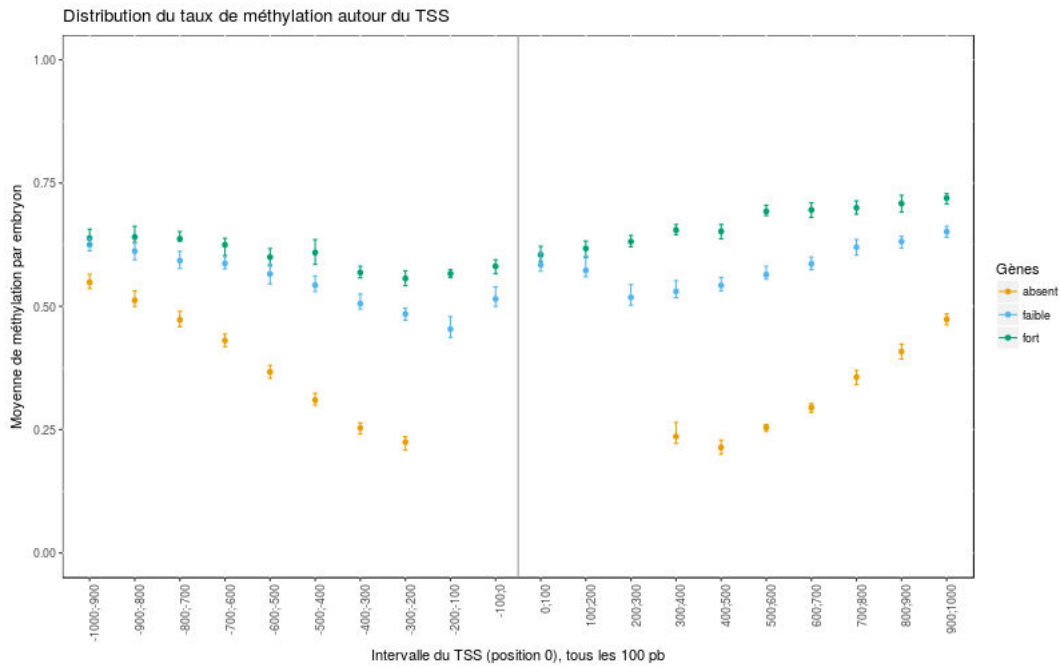
(a) Nombre de CG annotés avec la projection

(b) Nombre de CG annotés avec la redondance

FIGURE 3.9 – Nombre de CG dans les deux méthodes d'annotation. Les CG sont annotés dans les promoteurs, gènes et intergéniques selon les deux méthodes étudiées : projection et redondance.



(a) Distribution de la méthylation autour du TSS



(b) Distribution de la méthylation autour du TSS selon la densité des CG

FIGURE 3.10 – Taux de la méthylation autour du TSS. Chaque point représente une moyenne de méthylation par gène et par intervalle entre les 7 embryons (7 points par couleur et par intervalle). Chaque intervalle a une taille de 100 pb. Les points noirs sont la moyenne de méthylation de tous les CG détectés autour du TSS (a). Les points verts englobent les gènes avec des CG dans tous les intervalles ("fort") (b). Les points orange comprennent les gènes dont les CG sont absents entre -200 et +300 pb ("absent"). Et les points bleus sont tous les autres CG, c'est-à-dire il manque au moins un CG dans l'intervalle proche du TSS ("faible"). Les barres d'erreur représentent le taux de méthylation moyen le plus petit et le plus grand entre les 7 embryons, au sein d'un intervalle. La ligne grise représente la position du TSS.

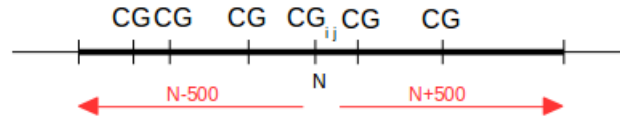


FIGURE 3.11 – Schéma de la fonction de lissage en différentiel de méthylation. Le schéma montre un exemple de site CpG positionné sur le génome à une position N , avec un taux de méthylation i , et qui appartient au groupe j . La fenêtre de lissage s'étend de ± 500 pb de part et d'autre du site.

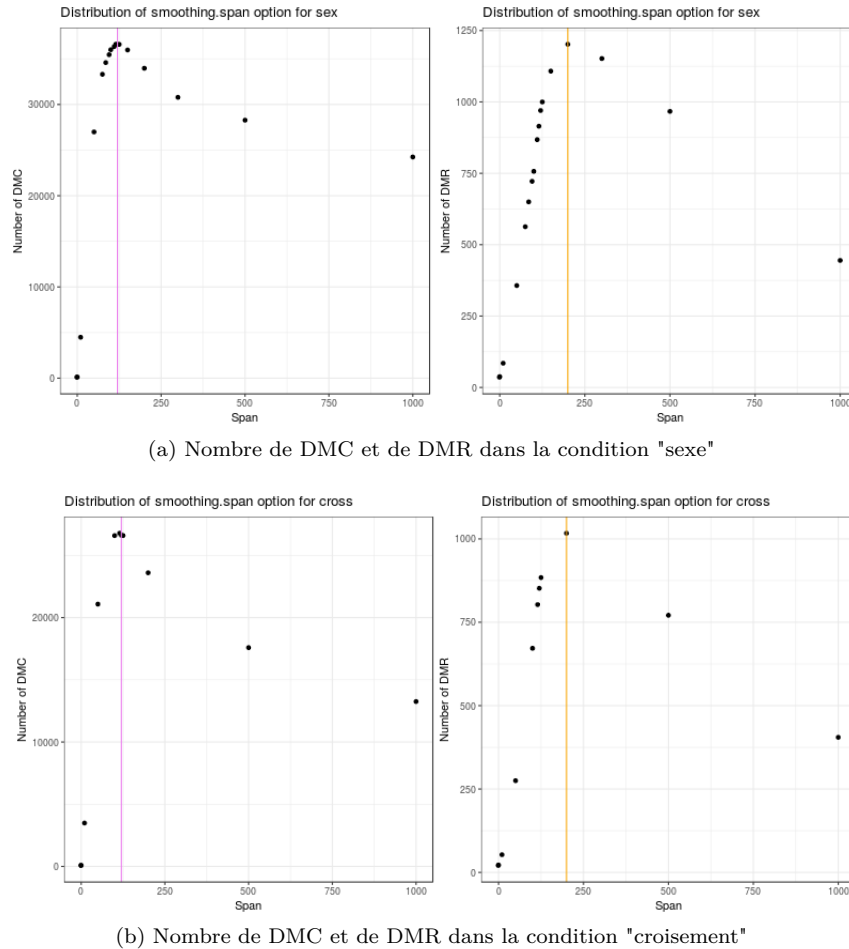


FIGURE 3.12 – Nombre de DMC et de DMR selon les fenêtres de lissage. Les 2 conditions de différentiels sont représentées. La ligne rose représente le maximum de DMC détectées, avec une fenêtre de 120 pb. La ligne orange révèle que le maximum de DMR détectées est avec une fenêtre de 200 pb.

Condition	Différentiel	Nb total	Sens UP	Sens DOWN
Sexe	DMC	36576	12414 (33,94%)	24162 (66,06%)
Sexe	DMR	970	406 (41,86%)	564 (58,14%)
Croisement	DMC	26615	10227 (38,43%)	16388 (61,57%)
Croisement	DMR	852	261 (30,63%)	591 (69,37%)

Tableau 3.5 – Nombre de DMC et de DMR pour chaque condition de différentiel. Le sens UP pour la condition "sexe" signifie que les femelles ont un CG sont plus méthylé que les mâles (DOWN : CG plus méthylé chez les mâles) tandis que pour la condition "croisement", le sens UP révèle un taux de méthylation du CG plus important pour le croisement $\sigma R^- \times \varphi 6$ (DOWN : croisement $\sigma 6 \times \varphi R^-$ est plus méthylé).

3.2.3.4 Paramètres pour le différentiel de méthylation

Après avoir caractérisé la répartition de la méthylation le long du génome pour tous les individus (cf projet de publication, partie 3.1 page 83), nous avons analysé la différence entre les taux de méthylation obtenus pour chaque groupe d'embryons : condition "sexe" (mâles et femelles) ou condition "croisement" ($R^- \times 6$ et $6 \times R^-$). Pour cela, la méthode du logiciel DSS se base sur une loi bêta-binomiale et un test statistique de Wald. Il calcule pour chaque cytosine au sein d'un ensemble d'individus (un sexe ou un sens de croisement) un taux de méthylation moyen puis compare ce taux au taux obtenu à la même position dans l'autre groupe testé (l'autre sexe ou l'autre sens de croisement). Cette cytosine qui présente un taux de méthylation différentiel est une DMC (*Differentially Methylated Cytosine*), lorsque le test de Wald est significatif avec une p-valeur inférieure à 5%. Sur nos données, nous avons choisi d'ajuster cette p-valeur avec un test de Benjamini-Hochberg (correction de tests multiples) [194] et ainsi obtenu une p-valeur ajustée, ou q-valeur, inférieure à 5% qui permet de déclarer le différentiel significatif et donc de détecter une DMC. De plus, dans les régions denses, les cytosines peuvent être regroupées le long du génome et former des ensembles appelés DMR (*Differentially Methylated Regions*) [195]. Ces régions génomiques sont définies par défaut par DSS comme étant d'une longueur minimale de 100 pb, au moins 3 CG y sont inclus, une p-valeur globale inférieure à 5%, 33% de CG dont la p-valeur est significative, une distance minimale avec la DMR voisine de 50 pb et au moins 10% de différence de méthylation entre les deux conditions. Ces taux de méthylation par condition sont calculés par la fonction de lissage sur une fenêtre de taille définie autour du CpG (Figure 3.11). Cette fenêtre de lissage est par défaut à 500 pb. Le package DSS a été développé sur des données humaines. Sur nos données, nous avons testé plusieurs valeurs pour détecter cette taille de fenêtre (Figure 3.12) : le nombre de DMC et de DMR est montré en fonction des conditions de différentiel : condition "sexe" (Figure 3.12 a), et condition "croisement" (Figure 3.12 b). Le maximum de DMC détectés sur les conditions "sexe" et "croisement" correspond à une fenêtre de 120 pb tandis que pour les DMR, il s'agit de 200 pb. Une DMR pouvant être définie par différents paramètres selon les études, nous avons choisi d'utiliser la fenêtre de lissage optimale pour la détection du nombre maximum de DMC, soit 120 pb. Les nombres de DMC et de DMR sont présentés dans le tableau 3.5 et dans le projet de publication (cf partie 3.1 page 83).

3.2.4 Méthylation allèle-spécifique

Ce chapitre reprend en partie les analyses décrites dans le projet de publication qui concernent la méthylation allèle-spécifique en lien avec l'expression allèle-spécifique, en discutant quelques éléments absents de la publication. Il élargit en outre les analyses en décrivant les résultats obtenus par croisement. On parle ici de phénomènes "allèle-spécifiques" quand les valeurs de méthylation ou d'expression entre les deux allèles sont significativement différentes (p valeur ajustée $< 0,05$).

Comme décrit dans l'introduction, le niveau de méthylation d'un dinucléotide CG peut dépendre de la séquence d'ADN à proximité de ce site. Afin de distinguer la méthylation de l'ADN entre deux allèles à partir de données de séquence, un polymorphisme de l'ADN doit être utilisé, comme un SNP. Les méthodes utilisées nous renseignent ainsi sur le niveau de méthylation et la séquence du même allèle. Ce type d'analyses repose donc sur la disponibilité de données de génotypage et de niveaux de méthylation dont les origines alléliques sont discernables, et donc dans l'idéal de données de séquençage génomique et de WGBS, ce qui est le cas dans notre dispositif. Comme toute analyse classique de niveau de méthylation par séquençage, la première étape consiste à aligner les lectures sur un génome de référence. Mais la présence de polymorphismes dans les sites informatifs peut poser le problème de biais d'alignement [196][197] : une lecture portant l'allèle de référence a plus de chances de s'aligner sur celui-ci qu'une lecture portant l'allèle alternatif, à cause d'un plus grand nombre de mésappariements (cf partie 1.3.3.1 page 67).

Une méthode d'analyse consiste à remplacer les positions de SNP connus dans le génome par des N [136][198][199], mais l'élimination des biais n'est pas totale [196]. Sur nos données, grâce au dispositif utilisé et à la disponibilité des génomes parentaux, nous avons fait le choix de construire une double séquence génomique ("double génome") contenant les haplotypes maternel et paternel [200]. Un haplotype est un ensemble d'allèles sur différents loci associés sur le même chromosome parental, généralement transmis ensemble à la génération suivante, aux événements de recombinaison près. Le double génome est donc construit en concaténant la séquence génomique maternelle d'une lignée et la séquence paternelle de l'autre lignée, élaborées à partir des données de séquence obtenues pour les 4 parents (cf Tableau 2.1 page 74). Il devient le génome de référence sur lequel les lectures provenant des embryons sont alignées. Aligner contre ce double génome permet d'obtenir

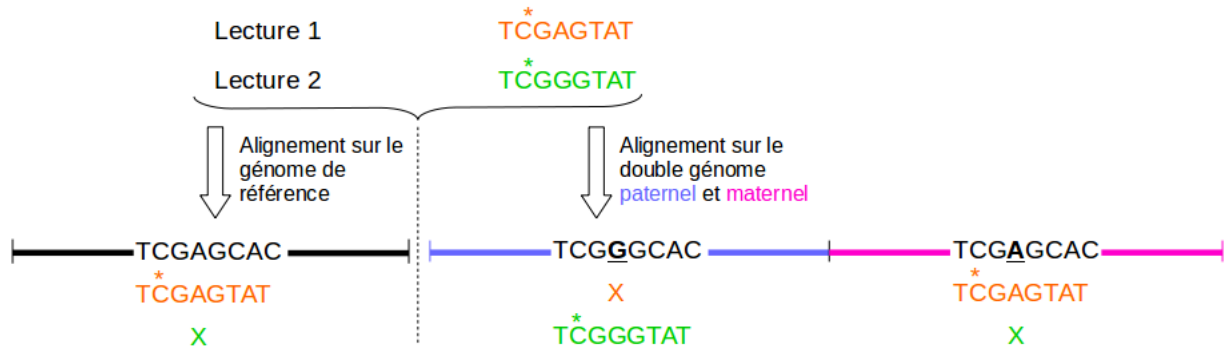


FIGURE 3.13 – Schéma de la méthylation allèle-spécifique détectée sur un double génome. Deux lectures d'un individu peuvent s'aligner sur le génome de référence et sur le double génome. Lorsque les lectures s'alignent sur le génome de référence, la lecture 1 orange s'aligne (obtention d'un taux de méthylation sur le CG représenté) tandis que la lecture 2 verte ne s'aligne pas à cause du SNP présent dans la lecture (base soulignée en noir). Lorsqu'on aligne les mêmes lectures sur le double génome maternel et paternel de l'individu, la lecture 1 orange s'aligne sur le génome maternel tandis que la lecture 2 verte s'aligne au même endroit sur le génome paternel. Un taux de méthylation est être calculé par origine parentale.

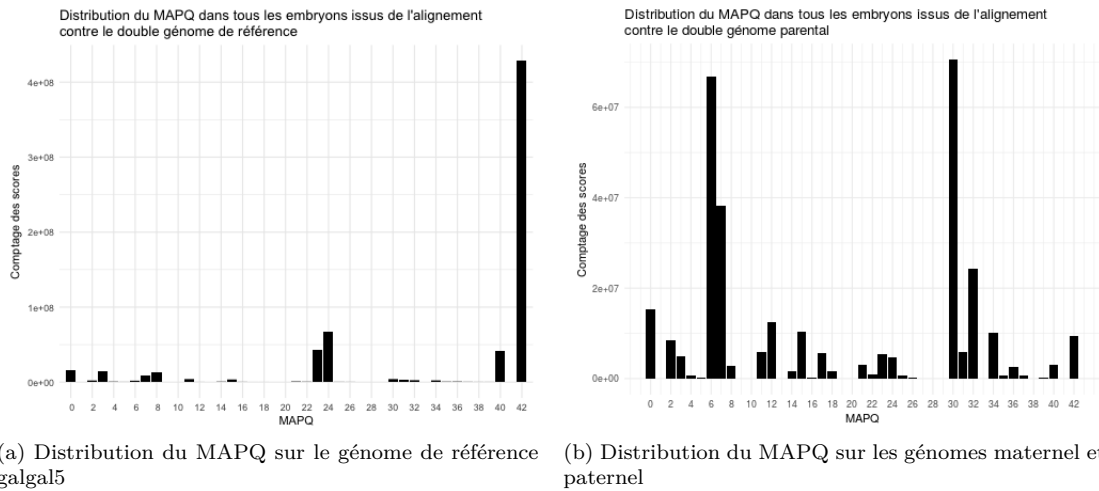


FIGURE 3.14 – Distribution du MAPQ sur le génome de référence (a) et sur le double génome (b). Le pic de distribution est à 42 sur les lectures qui s'alignent sur galgal5 (a), ce qui indique une forte probabilité que sa position est au bon endroit sur le génome. Deux pics de distribution ressortent lors de l'alignement des lectures sur le double génome (b) : MAPQ à 6 et MAPQ à 30. Cela indique la présence d'un SNP dans la lecture, ce qui induit un *mismatch* supplémentaire.

des positions de CG identiques retrouvées sur les deux génomes parentaux et de détecter une méthylation spécifique selon l'origine parentale (Figure 3.13).

En effet, dans une analyse de méthylation allèle spécifique, seules les lectures qui permettent de distinguer l'origine parentale sont informatives. Les autres peuvent être éliminées en filtrant leurs alignements. La stratégie du double génome repose sur la capacité à opérer ce filtre automatiquement sur la base de la valeur MAPQ reportée par l'aligneur, censée refléter la fiabilité de l'origine indiquée de la lecture. Une valeur de MAPQ positif indique que l'alignement avec le meilleur score ne peut venir que d'une seule position génomique (*unique best hit*). En outre, le MAPQ est d'autant plus élevé que l'écart de score avec le deuxième meilleur alignement est élevé. En réalité, les lectures provenant des régions dépourvues de polymorphisme peuvent s'aligner aussi bien sur l'haplotype paternel que maternel, avec le même nombre de *mismatches*. Les alignements résultants reçoivent donc une valeur MAPQ de 0 pour cause d'ambiguïté de position sur le double génome, et sont filtrés par le pipeline (étape de traitement par Bismark). Ne restent donc que les lectures qui peuvent "s'accrocher" de façon univoque, ou du moins préférable, sur un des deux haplotypes, permettant ainsi de cibler les positions d'intérêt pour l'ASM (Figure 3.13). Pour confirmer la qualité de l'alignement sur ce double génome, les MAPQ ont été observés (Figure 3.14) : d'abord lors de l'alignement contre le génome de référence *Gallus gallus* version 5, ou galgal5, puis lors de l'alignement contre ce double génome. Une majorité de lectures *paired-end* s'alignent avec un très haut score de qualité d'alignement (MAPQ = 42, Figure 3.14 a) tandis que des pics de distribution sont observés pour des MAPQ de 30 et de 6 (Figure 3.14 b).

Comme nous avons vu, une lecture s'aligne sur le double génome lorsqu'un SNP est présent et différencie le génome paternel du maternel. Ce *mismatch* fait baisser le score de qualité de l'alignement. La majorité des lectures avec un MAPQ de 42 lors de l'alignement sur galgal5 se retrouve avec un MAPQ de 30 et 6 sur le double génome (Figure 3.15) : dans la séquence seule une base a changé. La présence d'un variant dégénéré dans la lecture peut expliquer la différence entre les MAPQ de 30 et de 6. Le score de MAPQ reflétant l'écart entre le meilleur alignement et le second meilleur, la présence d'un variant dégénéré dont l'un des deux allèles est similaire à celui de l'autre génome parental, le score du second meilleur alignement est proche du meilleur alignement, ce qui entraîne une diminution du score à 6. Par exemple, dans le génome de référence, la séquence d'ADN est ACGAG, la séquence du génome maternel est la même, tandis que la séquence du génome paternel est ASGAG. Selon le code IUPAC (www.bioinformatics.org/sms/iupac.html), la base S représente soit un C soit un G. La lecture ACGAG de l'embryon s'aligne parfaitement sur le génome maternel mais peut très probablement s'aligner sur le génome paternel donc les deux meilleurs scores d'alignements sont proches, et le MAPQ diminue.

Les positions de CG sur le double génome sont détectées avec le même pipeline que précédemment, les paramètres restent inchangés (Tableau 3.6). Pour chaque individu, on obtient une liste de CG (position et taux de méthylation) pour les génomes maternels et paternels par individu. Le nombre de CG moyen détecté est de 6,99 millions pour les CG issus des parents 7127/7128 (haplotype lignée 6) et de 6,30 millions pour les CG issus des parents 8303/8403 (haplotype lignée R⁻) (Tableau 3.6), soit 39,07% et 35,23% de CG, respectivement dans les parents 7127/7128 et 8303/8403, par rapport au nombre de CG présents dans le génome de référence. On remarque que les parents 7127/7128 ont plus de CG détectés que les parents 8303/8403. Ceci est lié au plus grand nombre de lectures alignées sur les génomes 7127/7128 que sur les génomes 8303 et 8403. Même si les polymorphismes sont rares au sein de chaque lignée, les lignées ne sont pas totalement consanguines. Des variations au sein des haplotypes parentaux peuvent être présentes entre les embryons issus d'un même sens de croisement. La lignée 6, représentée par les parents 7127 et 7128, est très consanguine (0,04% de SNP hétérozygotes), de même pour la lignée R⁻, avec les parents 8303 et 8403, même si elle est légèrement moins homozygote (0,13% de SNP hétérozygotes) (Tableau 3.7). Comme attendu, les lignées sont très différentes l'une de l'autre. Cependant, le parent 8303 est un peu plus hétérozygote que le parent 8403. De la même manière, on observe plus de SNP hétérozygotes chez le parent 7128 (Tableau 3.7). On a donc maximisé le nombre de positions posant potentiellement des problèmes d'alignement dans ce croisement, ce qui peut expliquer la différence dans le nombre de CG détectés selon le sens de croisement : moins de lectures alignées dans le croisement 8303*7128 et moins de CG détectés.

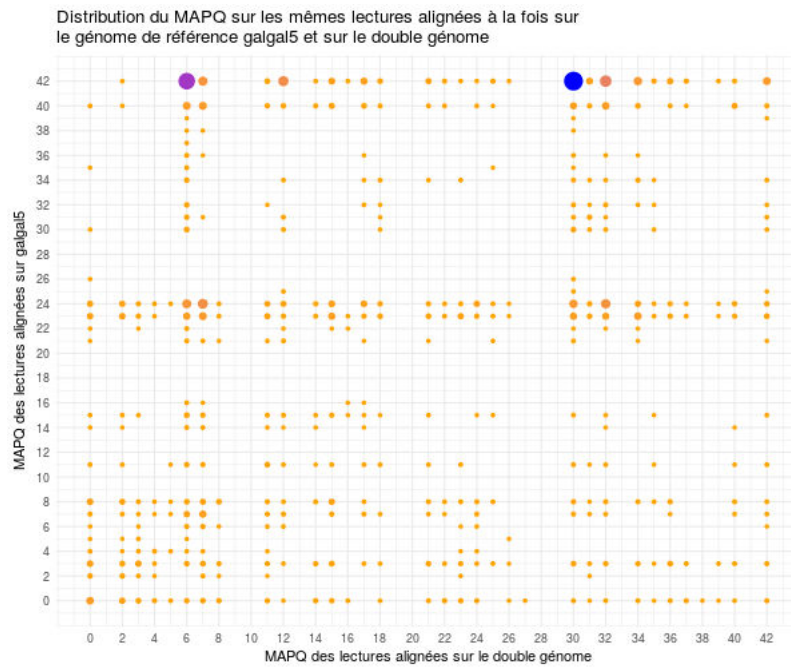


FIGURE 3.15 – Distribution du MAPQ des mêmes lectures alignées sur le génome de référence et sur le double génome. Sur les axes x et y est représentée la distribution du MAPQ des lectures alignées sur le double génome et galgal5 respectivement. L'intensité de couleur (de orange à bleu) et la taille des points indique le nombre de lectures : plus la couleur est foncée et plus le point est gros, plus il y a des lectures avec ce MAPQ. Les lectures avec un MAPQ à 42 sur galgal5 sont retrouvées dans les lectures avec un MAPQ à 30 (point bleu) et à 6 (point violet).

Embryons	Nb lectures	Nb CG	Profondeur CG (X)	% meth CG
7127emb08	18474270	6566650	3,93	70,19%
7127emb15	22436724	6633833	5,05	73,92%
7127emb20	28204040	6558248	4,45	71,43%
7127emb23	21691184	6786540	5,15	70,65%
8303emb12	31271692	7923621	5,85	70,74%
8303emb14	32324472	7806647	6,17	70,64%
8303emb22	20240378	6651988	4,43	71,64%
Moyenne	24948966	6989647	5,00	71,32%

(a) Génomes 7127/7128

Embryons	Nb lectures	Nb CG	Profondeur CG (X)	% meth CG
7127emb08	16211564	6197924	3,86	70,93%
7127emb15	19020588	6022144	5,17	74,69%
7127emb20	22964458	5974998	4,11	71,90%
7127emb23	19260360	6421301	5,18	71,36%
8303emb12	24310312	6955677	5,50	71,24%
8303emb14	24688034	6822070	5,70	71,29%
8303emb22	15874452	5718141	4,34	72,18%
Moyenne	20332824	6301751	4,84	71,94%

(b) Génomes 8303/8403

Tableau 3.6 – Données alignées sur le double génome. Le tableau a représente le nombre de lectures après alignement, le nombre de CG détectés, leur profondeur moyenne et le taux de méthylation moyen, des lectures des embryons provenant des parents 7127 et 7128 (lignée 6). Le tableau b, décrit les mêmes statistiques pour les lectures des embryons provenant des parents 8303 et 8403 (lignée R⁻).

Parents	Nb SNP homozygotes	Nb SNP hétérozygotes	% SNP hétérozygotes
7127	1217991675	500858	0,04%
7128	1217963684	528849	0,04%
8303	1216875082	1617451	0,13%
8403	1216945454	1547079	0,13%

Tableau 3.7 – Nombre de SNP des génomes parentaux. Le nombre de SNP homozygotes au sein de chaque parent est plus important que le nombre de SNP hétérozygotes. Globalement, peu de SNP hétérozygotes sont observés chez les 4 parents (moins de 1%) mais la lignée des parents 8303/8403 a un nombre légèrement plus élevé de SNP hétérozygotes.

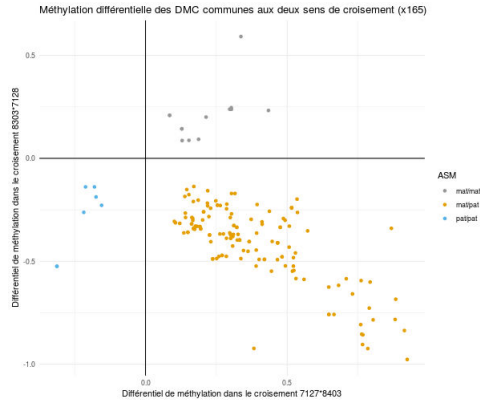


FIGURE 3.16 – Méthylation différentielle des DMC communes aux deux sens de croisement. La distribution de la méthylation différentielle concernent les 165 DMC provenant des deux sens de croisement. Les points gris et les points bleus représentent la méthylation différentielle des DMC portée par le même parent quelque soit sa lignée : en bleu, par le génome paternel (soit 7128 et 8403) et en gris, par le génome maternel (soit 7127 et 8303). En orange, ce sont les DMC qui ont une méthylation plus importante sur le parent issu de la même lignée dans les deux groupes (soit 7127 et 7128, soit 8303 et 8403).

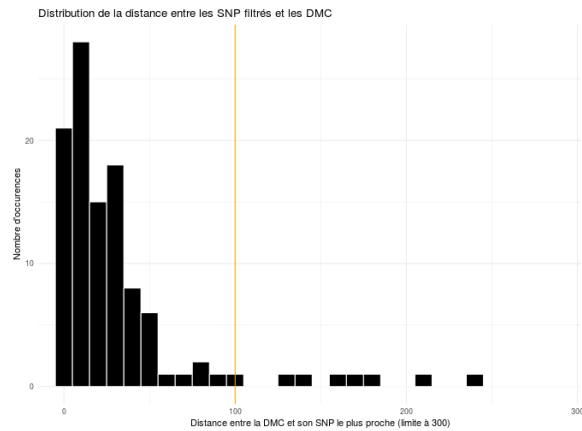


FIGURE 3.17 – Distance entre les SNP et les DMC. Cet histogramme représente la distribution de la distance la plus proche entre les SNP homozygotes par lignée et les DMC communes au deux sens de croisement. La majorité des distances est inférieure à 70 pb (chaque barre représente une distance de 10 pb). Pour obtenir le maximum de SNP proche d'une DMC, un seuil à 100 pb (taille de la lecture) a été choisi (ligne orange).

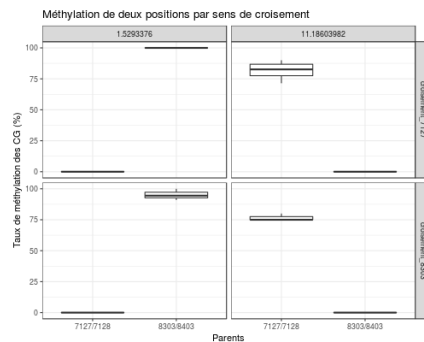


FIGURE 3.18 – Méthylation par allèle selon les sens de croisements. Il s'agit d'un exemple de deux positions de DMC associées à un SNP dans les croisements 7127 et 8303. Le résultat montre que, pour chaque lignée, un allèle est plus méthylé que l'autre. Le CG à la position 1.5293376 (chromosome 1) est plus méthylé sur le génome paternel dans le croisement 7127, c'est-à-dire sur 8403, et est plus méthylé sur le génome maternel dans le croisement 8303, c'est-à-dire sur 8303. Cela confirme une ASM.

3.2.4.1 ASM par croisement

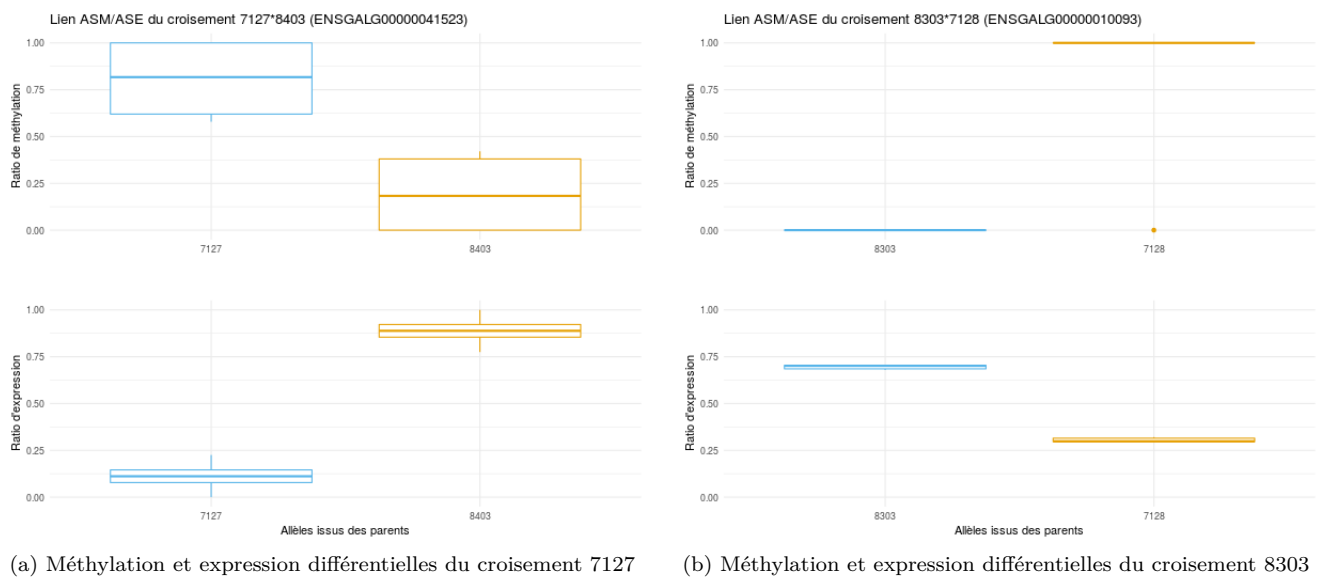
À partir des CG paternels et maternels, plusieurs jeux de données sont construits pour détecter l'ASM selon plusieurs conditions : par croisement et par haplotype. La première analyse détecte l'ASM au sein des deux croisements : on analyse indépendamment les 4 individus issus du croisement 7127 et les 3 autres issus du croisement 8303. Des filtres de couverture ont été appliqués comme pour les précédentes analyses (positions communes à tous et au moins 75% des individus ont un minimum de couverture à 5X). Dans le croisement 7127, pour 4 embryons, 627128 positions de CG sont analysées à la fois sur le génome maternel et paternel et 2440 DMC sont obtenues. Parmi elles, 1172 (48,03%) CG sont plus méthylés sur le génome maternel 7127 (lignée 6) et 1268 (51,97%) sur le génome paternel 8403 (lignée R⁻). Parallèlement, dans le croisement 8303, pour 3 embryons, 1664286 de CG présents à la fois sur les génomes paternel et maternel ont été analysés. Nous obtenons 7160 DMC dont 3900 (54,47%) avec une méthylation plus importante pour le génome maternel 8303 (lignée R⁻) et 3260 (45,53%) sur le génome paternel 7128 (lignée 6). Ces DMC sont toutes différentielles dans le même sens pour tous les embryons (la lignée R⁻ toujours plus méthylée que la lignée 6, ou l'inverse, quelque soit le croisement).

Ces deux groupes de DMC ont été comparés et 165 positions différentielles communes entre les deux sens de croisement s'en dégagent. Parmi ces 165, certaines (n = 21) ont un différentiel dans le même sens par rapport à l'origine parentale : le génome provenant de la mère (ou du père) est toujours plus méthylé(e) quelque soit le sens de croisement, malgré le filtre appliqué précédemment. Ce phénomène est caractéristique de l'empreinte génomique parentale. Comme l'absence de ce phénomène a été démontrée sur ce dispositif [131], une analyse approfondie de ces résultats est en cours.

Néanmoins, la majorité (n = 144) ont un différentiel de méthylation lié à la lignée. Par exemple, la même DMC est plus méthylée sur le chromosome provenant de la lignée 6, donc sur l'haplotype issu de la mère 7127, dans le croisement 7127, et sur l'haplotype issu du père 7128, dans le croisement 8303. Une méthylation différentielle est obtenues pour 165 DMC communes (Figure 3.16). Ces CG différentiels sont associés à un SNP à proximité sur la séquence. Pour tenter d'interpréter ce lien SNP / méthylation allèle-spécifique, nous avons observé les SNP situés à une distance inférieure à 100 pb du DMC analysé (taille de la lecture séquencée) : c'est dans cet intervalle que se situe le plus grand nombre de SNP associés à une DMC (Figure 3.17). Les SNP utilisés ont été détectés à partir des génomes parentaux et sont donc spécifiques de lignées. Afin de détecter l'origine parentale du SNP, ils doivent être homozygotes pour un allèle chez les parents 7127/7128 et homozygotes pour l'autre allèle chez les parents 8303/8403. De plus, un vrai SNP C/T ne peut pas être distingué d'une substitution C/T causée par la conversion au bisulfite et donc générer un taux de méthylation erroné (cf partie 1.3.3.1 page 67). On supprime donc ces SNP C/T (et G/A provenant du brin anti-sens). Au final, sur les 144 DMC ayant un différentiel lié à la lignée, 96 couples DMC-SNP sont obtenus après les différents filtres de SNP. Deux exemples de DMC parmi ces 96 sont représentées en fonction de l'origine parentale (Figure 3.18).

Sur ce total de CG différentiels, 7 DMC sont dans les promoteurs et 35 dans le corps du gène. Les 35 DMC sont associées à 27 gènes, dont 11 gènes ont un taux de méthylation plus important dans la lignée R⁻ et 16 dans la lignée 6. La majorité des 11 gènes associés à la lignée R⁻ sont impliqués dans le métabolisme, notamment dans les régions gouvernant la croissance et la taille. Il est probable que la sélection sur la consommation alimentaire résiduelle, qui a profondément impacté le métabolisme lipidique de la lignée R⁻ [201], ait affecté la régulation de l'expression de ces gènes. Parmi les 16 gènes associés à une méthylation plus importante dans la lignée 6, une partie de ces gènes sont impliqués dans le système immunitaire. La lignée 6 ayant été sélectionnée pour une meilleure résistance à des pathogènes [178], là encore il n'est pas exclu que des allèles particuliers intervenant dans la régulation de l'expression de ce gène aient été sélectionnés dans cette lignée. En outre, les 9 DMC présentes dans l'intervalle promoteur de 1kb sont associées aux gènes ENSGALG00000041523 (2 DMC), BRMS1L (*Breast Cancer Metastasis-Suppressor 1 Like*) (3 DMC) et ENSGALG00000027010 (4 DMC). BRMS1L code pour une composante d'un complexe de désacétylation des histones, et il a été proposé qu'il pourrait intervenir dans le remodelage de la chromatine en réponse au stress induit par une infection [202]. Comme précédemment, l'hypothèse la plus probable est que ce gène soit régulé par des allèles sélectionnés sur la résistance aux maladies. Les deux autres gènes n'ont pas été annotés.

En parallèle du travail de thèse sur les données de méthylation, une analyse similaire a été réalisée à partir



(a) Méthylation et expression différentielles du croisement 7127

(b) Méthylation et expression différentielles du croisement 8303

FIGURE 3.19 – Méthylation et de l'expression différentielles au sein de chaque croisement. La figure représente les valeurs de méthylation et d'expression de gènes ENSGALG00000041523 dans le croisement 7127*8403 (a) et BRMS1L dans le croisement 8303*7128 (b). L'axe y représente le ratio (valeur allèle / (valeur allèle A + valeur allèle B)) de la méthylation ou de l'expression pour chaque allèle. L'axe x représente le parent issu du croisement (lignée 6 en bleu et lignée R⁻ en orange).

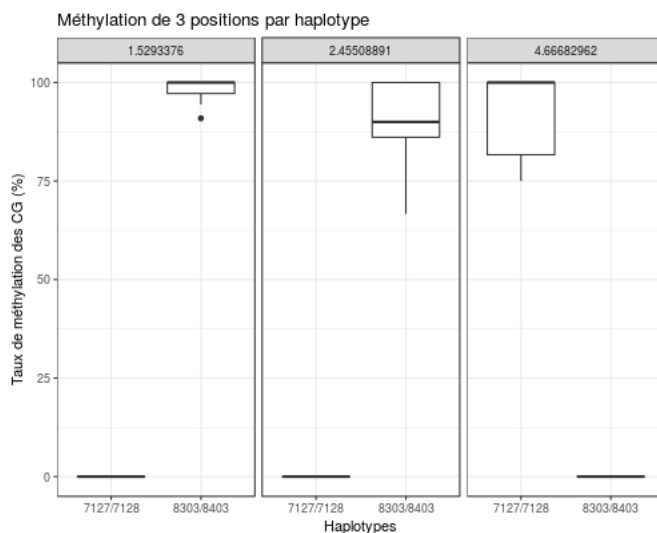


FIGURE 3.20 – Méthylation par allèle. Il s'agit d'un exemple de 3 DMC issues du différentiel par haplotype. Pour chaque lignée, un allèle est plus méthylé qu'un autre. Chaque boxplot représente 7 valeurs qui correspondent à la méthylation des 7 embryons dans une lignée.

des données de transcriptome, et un niveau d'expression a été calculé pour chaque allèle. Chaque lecture issue de RNA-seq a été alignée contre le même double génome constitué des deux parents pour pouvoir comparer les résultats d'ASM et d'ASE (cf projet de publication 3.1 page 83). Au final, 2209 gènes ont une expression différentielle selon l'allèle au sein du croisement 7127, et 1320 au sein du croisement 8303 (p-valeur ajustée < 0,05). Grâce à ces données d'expression allèle spécifique, il est possible de faire le lien avec la méthylation de ce même allèle dans les promoteurs. Un lien ASM/ASE est défini comme une différence significative observée à la fois entre l'expression des deux allèles d'un gène et entre les niveaux de méthylation des allèles du même gène. Seuls deux gènes montrent une expression différentielle en fonction de l'allèle associée à une DMC dans le promoteur. ENSGALG00000041523 révèle une ASE significative pour le croisement 7127 (p-valeur ajustée = 5.10^{-6}) et BRMS1L montre une ASE significative pour le croisement 8303 (p-valeur ajustée = 0,02). Lorsqu'on regarde le sens du différentiel entre l'ASE et l'ASM du même gène, on voit que le sens du différentiel de méthylation est à l'inverse du sens de différentiel sur l'expression (Figure 3.19). Dans le croisement 7127, l'allèle issu du parent 7127 est plus méthylé et moins exprimé que celui du parent 8303 au sein du croisement 7127 (Figure 3.19 a). Au contraire, au sein du croisement 8303, l'allèle issu du parent 7127 est moins méthylé et plus exprimé que celui issu du parent 8303 (Figure 3.19 b). La méthylation spécifique d'un allèle semble donc jouer un rôle dans la régulation de l'expression de ce même allèle, dans le sens d'une augmentation de l'expression associée à une diminution de la méthylation du promoteur.

3.2.4.2 ASM par haplotype

L'analyse précédente, par croisement, permettait de mettre en évidence des phénomènes potentiellement dépendants du sens de croisement. En effet, un allèle reçu par le père dans un sens de croisement peut ne pas se comporter comme le même allèle, reçu par la mère, dans l'autre sens de croisement. Dans cette partie, nous avons souhaité gagner en puissance et avons réalisé une analyse par haplotype. Les 7 embryons sont analysés ensemble : les CG sont regroupés par lignée, c'est-à-dire les CG issus des parents 7127 et 7128 dans un groupe (haplotype 7127) et les CG issus des parents 8303 et 8404 dans l'autre groupe (haplotype 8303) (cf Tableau 3.6 page 123). Les mêmes filtres et analyses que pour l'ASM par croisement ont été appliqués à ces deux groupes. Nous avons détecté 1325 DMC entre les deux haplotypes. Parmi ces 1325 DMC, 143 DMC sont retrouvées parmi les DMC communes analysées par croisement, dont le taux de méthylation est toujours plus important sur un haplotype.

Les filtres de SNP décrits précédemment sont appliqués à ces 1325 DMC pour obtenir 824 DMC associées à un SNP spécifique de lignée (pour lequel les deux lignées sont homozygotes, pour des allèles différents). Trois exemples sont illustrés sur la figure 3.20. La méthylation de ces trois DMC montre qu'un allèle est toujours très méthylé, ce qui confirme une ASM. On a ainsi plus de DMC associées à un différentiel selon les haplotypes que dans la première analyse par croisement. Parmi les 824 DMC, 31 DMC sont dans les promoteurs et 397 dans le corps du gène. Dans les 397 DMC, 2 sont annotées 5'UTR, 9 sont 3'UTR, 15 sont CDS et 28 dans les exons. Aucun enrichissement n'a été détecté. Les analyses d'enrichissement d'annotations ont pour but de calculer une probabilité : sachant qu'un groupe de gènes est annoté avec un terme spécifique (ontologie de gènes ou *gene ontology*), quelle est la probabilité que la totalité ou une fraction des gènes issus du jeu de données étudiés soient dans un groupe ciblé. Au sein des DMC liées au CDS ou aux promoteurs, quelques fonctions ressortent comme l'activité dans le métabolisme des lipides pour les DMC régulées dans la lignée R⁻, ce qui suggère comme précédemment une régulation de l'expression des gènes impliqués dans la consommation alimentaire résiduelle (Tableau 3.8). On observe également, comme précédemment, des fonctions liées à la résistance aux maladies.

Au sein des gènes associés aux 31 DMC annotées dans le promoteur 1kb, 23 gènes sont liés à une DMC, 2 gènes à 2 DMC (dont ENSGALG00000041523) et 2 gènes à 3 DMC (BRMS1L et ENSGALG00000027010). Ces trois gènes ont été détectés comme différentiellement méthylés et exprimés lors de l'analyse par croisement. Pour faire le lien de l'ASM par haplotype avec l'expression allèle-spécifique, un niveau d'expression par allèle a été calculé à partir des mêmes données de transcriptome que celles utilisées dans l'ASE par croisement. Dans l'analyse par haplotype, 1073 gènes ont montré une expression allèle-spécifique avec une p-valeur ajustée inférieure à 0,01. Deux gènes présentent à la fois un différentiel de méthylation dans le promoteur et un différentiel d'expression : ENSGALG00000041523 et ISPD (*Isoprenoid Synthase Domain Containing*). Le gène ISPD est impliqué dans la glycosylation, phénomène moléculaire qui ajoute des résidus de sucres à certaines protéines, et est notamment impliqué dans le fonctionnement de la protéine α -dystroglycan. Au niveau du

Tableau 3.8 – Principales fonctions des gènes associés aux DMC par haplotype. Pour les DMC liées aux CDS des gènes et aux promoteurs, le nombre de DMC, le nom complet du gène, ses principales fonctions dans les voies de régulation moléculaires et cellulaires ainsi que son annotation dans les caractéristiques géniques sont décrits pour chaque gène annoté.

Nom du gène (Ensembl)	Nb de DMC	Lignée la plus méthylée	Nom du gène annoté	Nom complet du gène annoté	Fonctions	Annotation
ENSGALG0000001926	1	6	HSPB1	Heat Shock Protein Family B (Small) Member 1	Métastases et croissance de cellules cancéreuses	CDS
ENSGALG00000005195	1	6	OSBPL2	Oxysterol Binding Protein Like 2	Récepteurs de lipides intracellulaires	CDS
ENSGALG00000002759	1	6	PEAK1	Pseudopodium Enriched Atypical Kinase 1	Métastases et croissance de cellules cancéreuses	CDS
ENSGALG00000030075	1	6				CDS
ENSGALG00000034349	1	6				CDS
ENSGALG00000025882	1	R-	BUD13	BUD13 Homolog	Mesure de triglycérides	CDS
ENSGALG00000006632	1	R-	CEP63	Centrosomal Protein 63	Métabolisme, homéostasie	CDS
ENSGALG00000040896	1	R-	FASN	Fatty Acid Synthase	Métabolisme et synthèse des acides gras	CDS
ENSGALG00000032840	2	R-	LYST	Lysosomal Trafficking Regulator	Fixation de phospholipides	CDS
ENSGALG00000009601	1	R-	NELL2	Neural EGFL Like 2	Système nerveux	CDS
ENSGALG00000037506	1	R-	SIPA1L2	Signal Induced Proliferation Associated 1 Like 2	Activation des GTPases	CDS
ENSGALG00000006192	1	R-	USH1C	USH1 Protein Network Component Harmonin	Différenciation de cellules rétinienne	CDS
ENSGALG00000044592	1	R-				CDS
ENSGALG00000045035	1	R-				CDS
ENSGALG00000007395	1	6	ABCC2	ATP Binding Cassette Subfamily C Member 2	Résistance aux médicaments	Promoteur
ENSGALG00000002198	1	6	FHOD1	Formin Homology 2 Domain Containing 1	Mesure de taux lipidiques	Promoteur
ENSGALG00000010795	1	6	ISPD	Isoprenoid Synthase Domain Containing	Développement embryonnaire	Promoteur
ENSGALG000000026572	1	6	RGS5	Regulator Of G Protein Signaling 5	Augmentation de la pression sanguine	Promoteur
ENSGALG00000036915	1	6	SQLE	Squalene Epoxidase	Biosynthèse du cholestérol	Promoteur
ENSGALG00000008507	1	6	THOC2	THO complex 2	Système nerveux	Promoteur
ENSGALG00000009815	1	6	TMEM5	Transmembrane Protein 5	Dystrophie musculaire	Promoteur
ENSGALG00000009169	1	6	TRMT6	tRNA Methyltransferase 6	Méthyltransférases	Promoteur
ENSGALG00000037698	1	6	ZFAT	Zinc Finger And AT-Hook Domain Containing	Réponse immunitaire	Promoteur
ENSGALG000000027010	3	6				Promoteur
ENSGALG00000029182	1	6				Promoteur
ENSGALG00000031752	1	6				Promoteur
ENSGALG00000041523	2	6				Promoteur
ENSGALG00000046152	2	6				Promoteur
ENSGALG00000036160	1	R-	BOLL	Boule Homolog, RNA Binding Protein	Développement cellules germinales	Promoteur
ENSGALG00000010093	3	R-	BRMS1L	Breast Cancer Metastasis-Suppressor 1 Like	Migration et invasion des cellules cancéreuses	Promoteur
ENSGALG00000004759	1	R-	DHRS7B	Dehydrogenase/Reductase 7B	Phénotype alimentaire/digestif chez la souris	Promoteur
ENSGALG00000034768	1	R-	ESF1	ESF1 nucleolar pre-rRNA processing protein homolog	Développement embryonnaire	Promoteur
ENSGALG00000006010	1	R-	FAM103A1	RNMT Activating MRNA Cap Methyltransferase Subunit	Méthyltransférases	Promoteur
ENSGALG00000014170	1	R-	MOGAT2	Monoacylglycerol O-Acyltransferase 2	Synthèse des acides gras	Promoteur
ENSGALG00000017823	1	R-	RF00093		ARN non codant	Promoteur
ENSGALG00000017822	1	R-	RF00138		ARN non codant	Promoteur
ENSGALG00000000721	1	R-	SLC41A1	Solute Carrier Family 41 Member 1	Transport transmembranaire de cations	Promoteur
ENSGALG00000024260	1	R-				Promoteur
ENSGALG00000032653	1	R-				Promoteur
ENSGALG00000037911	1	R-				Promoteur
ENSGALG00000043407	1	R-				Promoteur

cerveau, cette protéine aide à diriger la migration des neurones durant le développement embryonnaire [203]. En outre, le gène `ENSGALG00000041523` est également retrouvé en ASM par croisement (croisement 7127). Néanmoins, le gène `BRMS1L` révélé dans le croisement 8303 n'est pas présent dans la liste de gènes ASE. Il présente toutefois une expression allèle-spécifique proche de la significativité.

Les deux gènes obtenus ont un allèle issu de la lignée 7127 plus méthylé et moins exprimé que l'allèle issu de la lignée 8303 (Figure 3.21). Afin d'observer une tendance globale, ces analyses ont été réalisées pour chaque gène tous embryons confondus. Si on regarde le lien ASM/ASE pour chaque embryon, la méthylation allèle spécifique est corrélée négativement avec l'expression allèle spécifique (Figure 3.22). Ainsi, sur nos données, des gènes ont une méthylation allèle spécifique dans leur promoteur liés à une expression allèle spécifique, suggérant un rôle de régulation. Peu de gènes annotés ont été révélés. Cependant, un gène ayant un lien ASM/ASE code pour une protéine : `ISPD` est impliqué dans le développement embryonnaire, tandis que l'autre gène, `ENSGALG00000041523`, code pour un ARN long non codant, suggérant une régulation épigénétique de l'expression de ces ARN. Ce lien ASM/ASE a été démontré dans plusieurs espèces, des insectes à l'Homme [138][204], mais n'avait pas encore été observé chez le poulet, faute d'analyses d'ASM tout génome dans cette espèce.

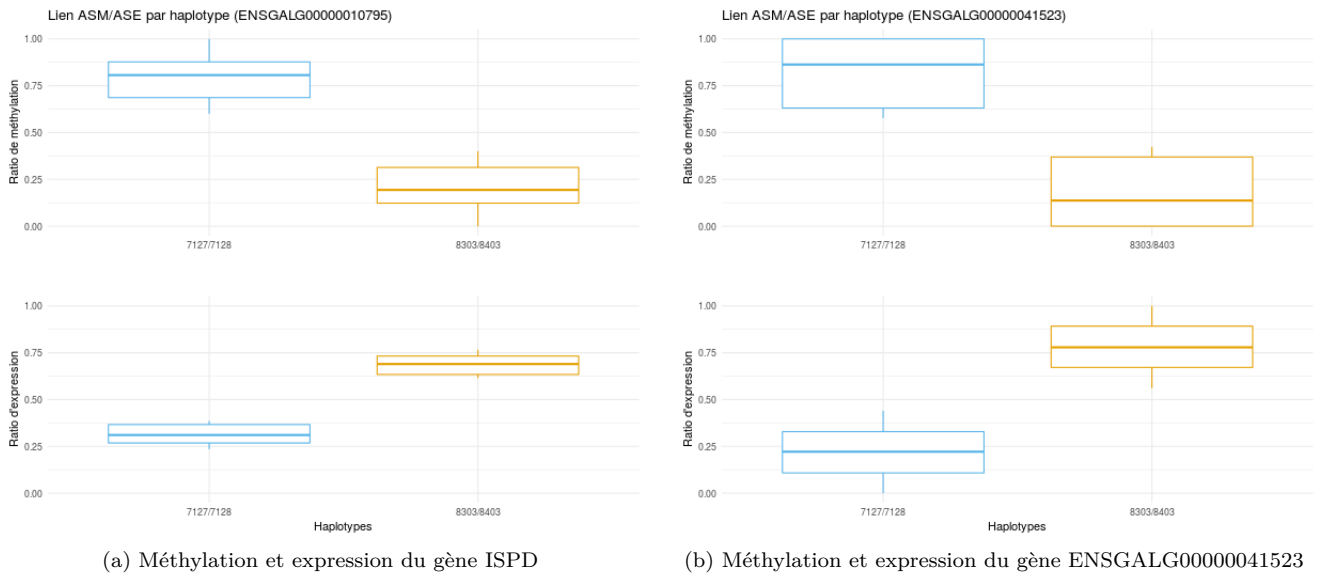


FIGURE 3.21 – Distribution de la méthylation et de l'expression différentielles par haplotype. Chaque figure représente un gène au sein duquel un ratio de méthylation et d'expression sont associés par haplotype. L'axe x décrit les deux haplotypes : 7127 en bleu et 8303 en orange. Comme pour la figure 3.19 précédente, un ratio de méthylation et d'expression est calculé afin d'homogénéiser les unités.

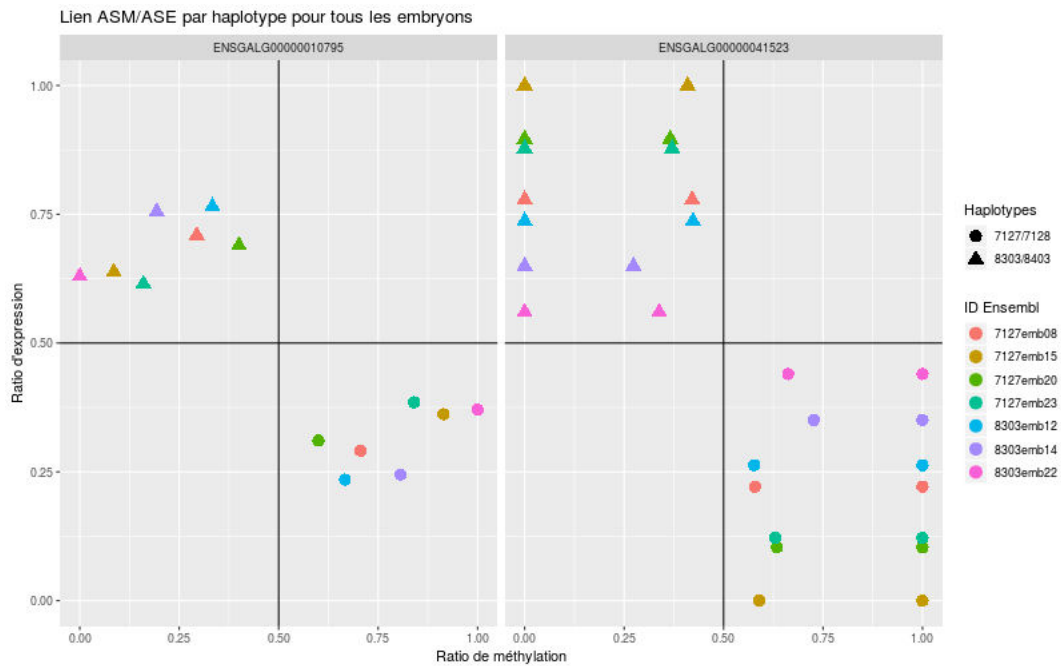


FIGURE 3.22 – Distribution de la méthylation et de l’expression allèles spécifiques dans tous les embryons ASM détectés dans le promoteur, associés aux valeurs d’ASE. Chaque couleur représente un gène avec un lien ASM/ASE et les deux formes, rond et triangle, concernent l’haplotype (rond pour l’haplotype 7127 et triangle pour l’haplotype 8303). Chaque point dessine le profil ASM/ASE pour un embryon dans un haplotype. Les axes noirs symbolisent un ratio de 50% pour l’expression et pour la méthylation. On considère que si la méthylation est supérieure à 50% et son expression correspondante inférieure à 50%, le lien ASM/ASE est corrélé de façon négative (plus c’est méthylé, moins c’est exprimé).

Chapitre 4

Analyses du méthylome en réponse à un stress environnemental

L'objectif de cette deuxième partie est d'élargir les analyses de la méthylation de l'ADN présentées dans le précédent chapitre à une autre question de recherche, en lien avec l'influence de l'environnement sur le méthylome. Ce projet permet d'étudier la réponse du méthylome chez la poule pondeuse à l'augmentation de la température dans l'environnement et à l'utilisation d'aliments peu énergétiques. Le dispositif construit dans ce projet contient également des données de génome et de transcriptome, ces dernières nous permettant de caractériser le lien entre la méthylation de l'ADN et l'expression des gènes. La mesure du taux de méthylation est effectuée sur plusieurs tissus adultes : foie, hypothalamus et sang total. Seules les données d'hypothalamus ont été générées à l'heure actuelle de la rédaction de la thèse et sont analysées dans ce chapitre. Les analyses présentées dans ce chapitre sont des analyses préliminaires nécessitant une recherche plus approfondie dans la littérature, en raison de l'arrivée récente des données. Les données sur le foie et le sang sont toujours en cours d'acquisition au moment de la rédaction de cette thèse, et la poursuite de ce travail est discutée dans le chapitre 5, partie 5.3 page 165.

4.1 Contrôle de la qualité des données RRBS

4.1.1 Des lectures de séquençage aux CG analysés

Le pipeline développé pour l'analyse des données WGBS et les analyses de méthylation allèle spécifique présentées dans le chapitre 3 a été utilisé sur les données de méthylome d'une sous-représentation du génome en réponse à un stress environnemental (données RRBS). Ce type de données a déjà été généré chez l'Homme [184], chez des organismes modèles [205][206][207], chez le porc [185] et chez la chèvre [208][186]. Chez le poulet, aucune étude d'analyses de données RRBS n'a été publiée à l'heure de la rédaction de cette thèse. Comme nous avons vu précédemment, le génome de référence *Gallus gallus* version 5 (GCA_000002315.3) a été digéré *in silico* par l'enzyme MspI afin de sélectionner des fragments dont la taille est comprise entre 40 et 100 pb (cf partie 2.2.1 page 75). L'ADN génomique a été extrait de l'hypothalamus de 36 individus adultes, pour obtenir des lectures *single-end* 1 x 50 pb. En moyenne, 21,49 millions de lectures sont obtenues (Tableau 4.1). Après suppression des séquences d'adaptateurs et filtres sur la qualité des séquences, les lectures sont alignées sur le génome de référence *Gallus gallus* version 5 (galgal5) donnant 34,35% de lectures finalement analysables par rapport aux lectures initiales (7,38 millions de lectures en moyenne après alignement). Les lectures ont été alignées de façon unique avec un taux d'alignement moyen à 48,56%.

Aucune étape de suppression de duplicats de PCR n'est réalisée. À l'issue de l'alignement, une moyenne de 1,44 millions de CG sont détectés dans un contexte CG avec une profondeur moyenne de 24,78X (Tableau 4.1). Quelques individus, notamment H_210_0592_Rm_FS, ont un nombre de lectures qui diminue drastiquement lors de la suppression des adaptateurs. Il s'agit d'une contamination de dimères d'amorces qui a eu lieu lors de la préparation des librairies. De plus, certains échantillons ont une faible profondeur. Nous choisissons d'éliminer les échantillons pour lesquels la profondeur moyenne est inférieure au minimum de couverture que nous allons appliquer dans la suite du pipeline, à savoir 12X. Les 4 échantillons supprimés sont repérés en

Échantillons	Nb lectures initiales	Nb lectures après alignement	Taux alignement	Nb CG	Profondeur CG	% meth CG
H_344_0780_Rm_HS	631273	89958	21,70%	295654	1,63	10,69%
H_094_1102_Rp_CT	5837160	123981	7,35%	359407	1,72	12,39%
H_103_0738_Rm_CT	9223034	1957766	33,35%	1057853	10,67	9,67%
H_009_1073_Rp_FS	65646612	2776992	33,80%	1159590	11,72	12,87%
H_375_0933_Rm_HS	6828390	3554197	53,30%	1231734	14,39	12,93%
H_342_0819_Rm_HS	8858601	3918440	49,75%	1348145	14,63	12,97%
H_006_0978_Rp_FS	12030773	4878985	46,45%	1359501	14,79	21,30%
H_427_0739_Rm_HS	17377923	3701371	35,80%	1137143	15,05	17,30%
H_204_0920_Rm_FS	9428495	4678129	53,40%	1410909	16,51	16,67%
H_210_0592_Rm_FS	82022872	4631234	29,60%	1407172	17,26	12,15%
H_341_1094_Rp_HS	11787914	6891094	58,90%	1589339	19,70	15,00%
H_446_0610_Rm_HS	10125774	5279908	55,05%	1189711	18,02	23,58%
H_007_0992_Rp_FS	25475322	4742533	39,95%	1179792	18,70	18,67%
H_357_0598_Rm_HS	10462478	5411084	55,05%	1284984	19,29	16,46%
H_417_1121_Rp_HS	12478152	6914813	56,30%	1476551	20,16	17,94%
H_004_0966_Rp_FS	12778364	7119653	56,45%	1511652	21,51	18,22%
H_208_0827_Rm_FS	17411960	6374031	48,75%	1475956	22,29	13,45%
H_051_0777_Rm_CT	22507119	8014109	49,45%	1712938	25,80	12,14%
H_202_0624_Rm_FS	20686775	8185005	55,90%	1698602	27,08	7,48%
H_470_1055_Rp_HS	30316603	6037891	40,05%	1167369	23,52	19,29%
H_014_1007_Rp_FS	17887871	8447633	55,65%	1644595	27,46	11,46%
H_463_1077_Rp_HS	18204752	7924389	55,10%	1538856	28,03	8,51%
H_410_0984_Rp_HS	16396727	10267894	64,00%	1805035	29,73	9,73%
H_161_0835_Rm_CT	18163944	9783244	56,10%	1799651	30,06	10,38%
H_099_0602_Rm_CT	20894306	10439662	58,10%	1839709	33,52	7,71%
H_108_1122_Rp_CT	24764689	10067735	48,25%	1726181	31,36	12,53%
H_197_0815_Rm_FS	21549156	11800199	59,20%	1800941	32,74	11,89%
H_062_1006_Rp_CT	49421181	10379851	57,65%	1823274	35,00	6,10%
H_065_0629_Rm_CT	27414157	11349829	47,15%	1817550	33,66	10,68%
H_203_0754_Rm_FS	21223614	10782619	54,35%	1484144	30,70	22,03%
H_185_1150_Rp_CT	24725118	11670606	49,75%	1772906	36,95	11,25%
H_168_1060_Rp_CT	24843549	12927931	55,40%	1927642	37,43	11,76%
H_132_0972_Rp_CT	20587056	9704756	52,15%	1530891	40,21	7,91%
H_013_1140_Rp_FS	25226706	11943834	50,95%	1536944	38,00	16,89%
H_088_0935_Rm_CT	24559139	12872896	56,35%	1595886	45,73	10,43%
H_331_1020_Rp_HS	25751999	10095040	47,50%	1109659	46,92	15,53%
Moyenne	21486932	7381647	48,56%	1439107	24,78	13,50%

Tableau 4.1 – Données issues de l’hypothalamus. Les lectures RRBS issues des hypothalamus sont alignées sur le génome de référence galgal5. Le nombre de lectures initiales est de 100% ; le nombre moyen des lectures restantes par rapport au nombre de lectures initiales après l’étape de nettoyage/filtres est de 78,95% ; et le nombre de lectures restantes après l’alignement est de 39,73%. Le taux d’alignement moyen des lectures des 36 individus est de 48,56%. Le nombre de CG détectés est de 1 439 107, pour une profondeur moyenne de 24,78X. La dernière colonne représente le taux de méthylation par individu, avec un taux moyen de 13,50%.

rouge dans le tableau 4.1. Nous observons alors un taux d’alignement moyen de 51,62% après les avoir éliminé. Pour chaque condition CT, HS et FS, le taux d’alignement moyen des lectures est de 53,04%, 51,89% et 50,06% respectivement. Plusieurs études ont montré que la taille des fragments sélectionnés et la longueur des lectures séquencées déterminent le taux d’alignement. Chez la chèvre, les auteurs montrent que des lectures de 2 x 125 pb s’alignent avec un taux d’alignement d’environ 64% : les fragments sélectionnés ont une taille de 250-500 pb [186]. Néanmoins une précédente analyse RRBS chez la chèvre révèle qu’à partir des fragments de 50-150 pb, des lectures de 2 x 100 pb ont un taux d’alignement de 38% et que l’augmentation de la taille des fragments à 150-250 pb permet d’augmenter le taux d’alignement à 61% [208]. Chez le porc, le taux d’alignement est d’environ 60%, avec des fragments de 110-220 pb et un séquençage de 2 x 100 pb [185]. Sur nos données, la digestion *in silico* sur le génome de référence et la connaissance du nombre de fragments obtenus par individu nous permet de définir des fragments de taille entre 40 et 100 pb, plus courts que dans les publications ci-dessus, ce qui peut expliquer en partie que le taux d’alignement de nos lectures (1 x 50 pb) est légèrement inférieur à ce qui a été observé. Cette discussion des résultats est reprise dans la partie 5.1.1 page 153.

Au final, le nombre de CG moyen sur les 32 individus est de 1,53 million avec une profondeur de 27,07X. Dans les trois conditions CT, HS, FS, le nombre de CG détectés est de 1,75 million, 1,35 million et 1,50 million respectivement (profondeur moyenne de 34,97X ; 22,67X ; 24,27X respectivement). À partir des 1,53 millions de CG, les différents filtres sont appliqués comme décrits précédemment. Le minimum de couverture est défini à 12X pour détecter un maximum de positions (individu le moins profond à 14,39X), sachant que la littérature recommande au minimum 10X de couverture pour le RRBS [208]. Parmi les positions communes entre les individus, ce minimum de couverture doit être présent dans au moins 75% des positions, soit dans 24 individus. De plus, un filtre sur les positions de SNP, issus des données génomiques obtenues sur les lignées R⁺ et R⁻ dans le cadre d’un autre programme (Sandrine Lagarrigue, communication personnelle), est utilisé pour supprimer les positions de CG ambiguës. Au final, 289 414 positions communes de CG filtrés pour chaque individu sont obtenues, avec une profondeur moyenne de 135X. Le niveau de méthylation dans les conditions CT, HS et FS est de 10,99%, 14,87% et 15,7% respectivement. Ces faibles niveaux de méthylation ont été également observés dans d’autres espèces, notamment chez l’Homme [174], chez le bovin [209], le cheval [210] et le porc [211].

4.1.2 Contrôle de la conversion au bisulfite

Comme pour les analyses tout génome, un taux de conversion des C hors CG est calculé en divisant le nombre de T par la couverture. Le taux moyen de conversion est de 98,48% sur les 32 individus (Tableau 4.2). Ce taux de conversion est correct, même si une meilleure qualité (>99%) aurait été préférée. Le même plasmide qu’en WGBS a été ajouté lors de la préparation de la librairie, non méthylé pour contrôler le taux de conversion ou méthylé pour contrôler le taux de surconversion. Son génome a également été ajouté au génome de référence pour les alignements. Le tableau 4.2 récapitule pour tous les individus, le type de plasmide ajouté et les taux de méthylation des C en contextes CpG, CHG et CHH. Les C hors CG sont très peu méthylés (1,09% pour CHG et 1,12% pour CHH), ce qui démontre une conversion correcte. Sur les 16 plasmides non méthylés, le taux moyen de méthylation des C en contexte CpG de 2,35%, soit un taux de conversion un peu trop faible, à 97,65%. Sur les 16 plasmides méthylés, le taux moyen de méthylation des C en contexte CpG est de 90,83%, indiquant un taux de surconversion d’environ 10%. Cela pourrait être dû à une trop importante dégradation lors de sa préparation. En effet, l’ADN du plasmide a été fragmenté de très nombreuses fois en amont de la préparation de la librairie, et nous soupçonnons une dégradation de l’ADN conduisant à un taux de conversion assez peu satisfaisant, et à un taux de surconversion trop important. Les résultats qui en découlent ne sont pas interprétables de manière certaine, et donc le plasmide ne peut pas servir de témoin pour la conversion au bisulfite. Néanmoins, le taux de conversion calculé sur les données génomiques elles-mêmes, proche de 100%, nous permet d’avoir confiance en la conversion sur nos données.

Échantillons	Taux de conversion	Plasmide	Contexte CpG	Contexte CHG	Contexte CHH
H_004_0966_Rp_FS	98,49%	Non méthylé	2,08%	1,10%	1,08%
H_006_0978_Rp_FS	98,48%	Méthylé	96,96%	2,37%	2,06%
H_007_0992_Rp_FS	98,46%	Méthylé	89,49%	1,21%	1,23%
H_013_1140_Rp_FS	98,60%	Non méthylé	2,07%	0,46%	0,83%
H_014_1007_Rp_FS	98,98%	Non méthylé	1,50%	0,29%	0,58%
H_051_0777_Rm_CT	98,41%	Non méthylé	2,81%	1,12%	1,09%
H_062_1006_Rp_CT	98,98%	Non méthylé	2,00%	0,90%	1,12%
H_065_0629_Rm_CT	98,48%	Non méthylé	2,69%	1,32%	1,46%
H_088_0935_Rm_CT	95,24%	Non méthylé	5,77%	1,04%	1,25%
H_099_0602_Rm_CT	98,67%	Méthylé	87,73%	1,20%	1,19%
H_108_1122_Rp_CT	98,76%	Non méthylé	1,76%	0,97%	1,00%
H_132_0972_Rp_CT	97,50%	Méthylé	85,68%	1,19%	1,73%
H_161_0835_Rm_CT	98,96%	Méthylé	90,18%	0,67%	0,77%
H_168_1060_Rp_CT	98,92%	Méthylé	87,87%	0,85%	0,71%
H_185_1150_Rp_CT	98,47%	Méthylé	87,93%	1,20%	0,98%
H_197_0815_Rm_FS	98,71%	Méthylé	96,68%	0,78%	1,15%
H_202_0624_Rm_FS	99,00%	Non méthylé	1,93%	0,47%	0,79%
H_203_0754_Rm_FS	98,50%	Non méthylé	2,95%	0,84%	0,95%
H_204_0920_Rm_FS	98,92%	Non méthylé	1,53%	0,48%	0,42%
H_208_0827_Rm_FS	98,89%	Méthylé	88,80%	0,90%	0,78%
H_210_0592_Rm_FS	96,67%	Méthylé	88,10%	4,67%	4,66%
H_331_1020_Rp_HS	97,73%	Non méthylé	2,60%	1,01%	1,18%
H_341_1094_Rp_HS	98,54%	Méthylé	97,33%	1,60%	1,12%
H_342_0819_Rm_HS	99,02%	Non méthylé	2,34%	0,81%	0,72%
H_357_0598_Rm_HS	98,94%	Non méthylé	1,63%	0,76%	0,68%
H_375_0933_Rm_HS	98,89%	Méthylé	89,87%	0,97%	0,77%
H_410_0984_Rp_HS	99,02%	Méthylé	90,83%	0,79%	0,79%
H_417_1121_Rp_HS	98,65%	Non méthylé	1,54%	1,24%	1,16%
H_427_0739_Rm_HS	98,77%	Méthylé	89,32%	0,75%	0,79%
H_446_0610_Rm_HS	98,53%	Méthylé	97,44%	1,23%	1,11%
H_463_1077_Rp_HS	98,80%	Méthylé	89,14%	0,92%	0,76%
H_470_1055_Rp_HS	98,32%	Non méthylé	2,41%	0,74%	0,98%
Moyenne	98,48%	50/50	90,83% / 2,35%	1,33% / 0,85%	1,29% / 0,96%

Tableau 4.2 – Contrôles de la conversion au bisulfite. Les taux de conversion des C non méthylés en T sont en moyenne à 98,48% sur les 32 individus. Le type de plasmide ainsi que les taux de méthylation des C dans les contextes CpG, CHG et CHH sont décrits dans ce tableau par individu. Les moyennes des taux de méthylation des 3 contextes sont données d'abord pour les plasmides méthylés, puis pour les plasmides non méthylés.

4.2 Méthylation des CG

4.2.1 Annotation des CG

Comme dans l'étude précédente, nous avons caractérisé la répartition des CG observés en fonction des régions fonctionnelles du génome. Pour annoter les CG, nous utilisons la méthode par projection détaillée ci-dessus (cf partie 3.2.3.2 page 113). Le promoteur est défini comme une région de 1 kb avant le TSS. Après filtres, 51,73% des CG sont retrouvés dans le gène et 18,31% dans le promoteur (Figure 4.1). En proportion, plus de régions promotrices sont détectées qu'en WGBS grâce au choix de l'enzyme MspI qui permet de cibler les régions riches en CG [159]. La méthylation est en moyenne de 5,6% dans les promoteurs, avec une majorité de CG entre 0% et 5% (Figure 4.2 a). Le taux de méthylation est plus élevé dans le gène (17,6% en moyenne), mais la majorité des CG restent non méthylés (Figure 4.2 b), ce qui est attendu, le RRBS ciblant les régions riches en CpG donc potentiellement hypométhylées [212]. Au sein du gène, la majorité des CG observés est située dans les introns, avec une méthylation moyenne de 14,8% (Figure 4.3). Il a été montré que la densité en CG des introns est plus importante que celle des régions non traduites (5'UTR et 3'UTR) et des exons [213]. De plus, les régions 5'UTR sont les régions les moins méthylées au sein du gène (4,3%) : elles sont proches du taux de méthylation moyen des promoteurs. La majorité des CG détectés sont retrouvés dans les îlots CpG qui sont principalement au sein des promoteurs (Figure 4.4). Plus on s'éloigne des CGI, plus le nombre de CG diminue, ce qui permet de valider la détection des CG plus importante dans les régions riches en CG ciblées par le RRBS. Les CGI non méthylés sont généralement associés à une activité transcriptionnelle de leur gène associé. Cependant, une partie des CGI a un taux de méthylation très contrasté avec ses régions voisines (*shores et shelves*). Ces CGI particuliers sont retrouvés près des gènes les plus exprimés, qui sont souvent des gènes de ménage chez l'Homme [214]. En conclusion, cette répartition de la méthylation le long du génome dessine un profil de méthylation avec des CG principalement non méthylés, dont une majorité sont dans les CGI. Le taux de méthylation moyen pour tous les échantillons dans les CGI est plus faible dans les régions promotrices qu'au sein des régions géniques et intergéniques (Figure 4.5). Cela confirme la présence de CGI, peu méthylés, dans les promoteurs [215].

Pour compléter avec l'annotation, comme dans les données WGBS, nous avons observé la méthylation autour du TSS (Figure 4.6). On observe que la méthylation diminue au niveau du TSS, ce qui est en accord avec la littérature (cf revue [85]). La plus grande profondeur du RRBS qu'en WGBS permet de mieux caractériser les régions riches en CG, notamment les CG autour du TSS. Le nombre de CG détectés est donc plus homogène autour du TSS, comparé aux données WGBS. C'est pour cela qu'aucun pic artéfactuel au TSS n'est observé. De plus, on observe que plus on s'éloigne du TSS, plus le taux de méthylation augmente, ce qui est attendu. Le taux de méthylation est plus faible à -2 kb qu'à 2 kb. En effet, la méthylation est plus importante dans le corps du gène qu'en amont de celui-ci (régions promotrices en amont du TSS).

4.2.2 Lien entre le taux de méthylation et le niveau d'expression

À partir des positions de CG filtrés, on analyse le lien entre la méthylation dans les promoteurs et l'expression des gènes associés. Les niveaux d'expression des gènes ont été détectés par RNA-seq et analysés par ailleurs. Brièvement, l'analyse RNA-seq est décrite à la suite. Les lectures *paired-end* ont été générées en 2 x 150 pb. En premier lieu, les séquences d'adaptateurs nécessaires au protocole expérimental sont supprimées des lectures par le logiciel Illumina CASAVA pour les analyses. Puis ces lectures ont été alignées sur le génome de référence *Gallus gallus* version 5 (GCA_000002315.3) en utilisant le logiciel STAR [216]. Les duplicats de PCR sont supprimés. Enfin, la quantification du niveau d'expression est réalisée avec le logiciel RSEM qui utilise l'annotation 87 de *Ensembl* [217]. Les valeurs d'expression des gènes sont estimées en TPM, comptage normalisé de transcrits par gène. Cette mesure donne une valeur d'expression relative comparable entre différentes conditions et entre différents gènes. Sur les 18821 gènes détectés en RNA-seq, 5916 gènes ont une valeur d'expression qui peut être associée à une mesure de méthylation dans le promoteur pour tous les individus. La répartition des niveaux d'expression se fait en quatre quantiles (Figure 4.7) : le niveau d'expression est symbolisé par des quantiles : le premier quantile est le moins exprimé et le quatrième quantile est le plus exprimé. On observe que la méthylation diminue au fur et à mesure des quantiles d'expression, ce qui est attendu. La distribution de la méthylation plus variable dans le premier quartile peut être due à la composition du quantile en gènes peu ou pas exprimés. Cette faible expression peut être corrélée à une méthylation plus forte du promoteur, mais également, à une spécificité tissulaire dépendante, par exemple, de la présence ou non d'un facteur de transcription.

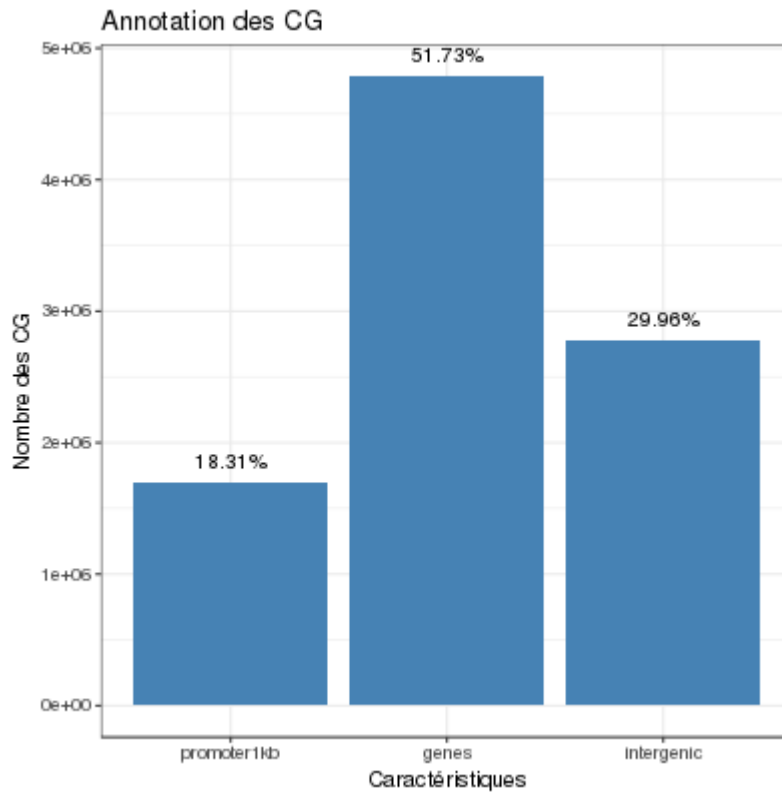
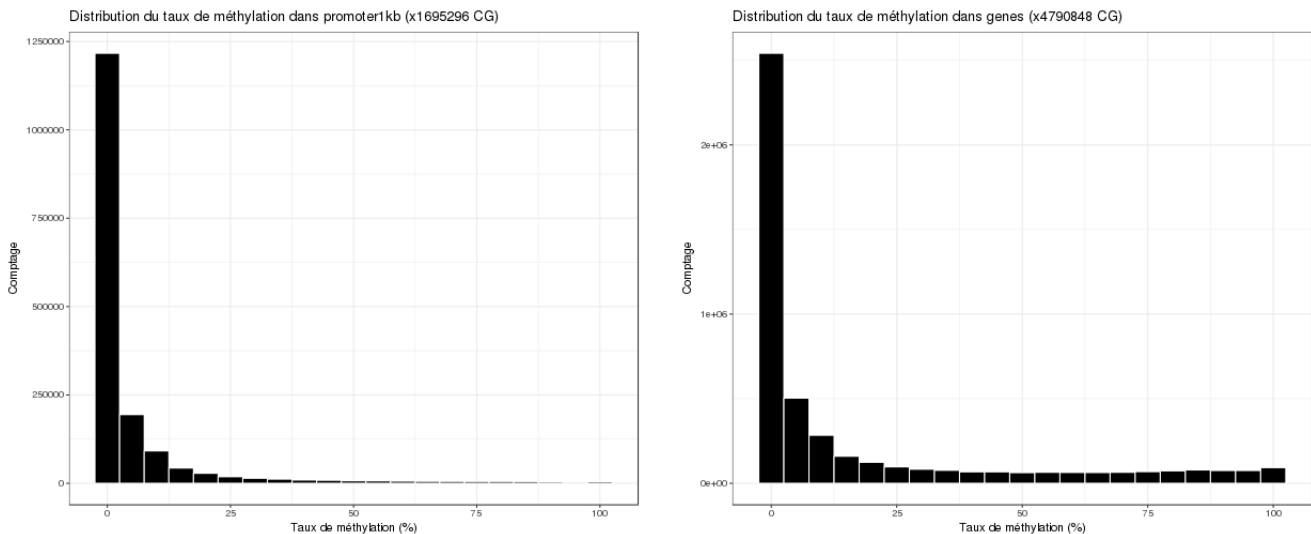


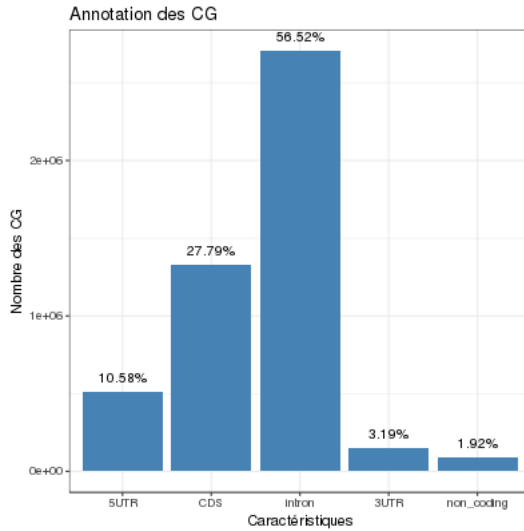
FIGURE 4.1 – Distribution du nombre de CG dans les caractéristiques génomiques pour tous les individus. Les caractéristiques génomiques sont les promoteurs, les gènes et les régions intergéniques. Le total du nombre de CG dans chaque intervalle est représenté sur l'axe y. Le pourcentage de CG trouvés pour chaque caractéristique est noté sur les barres de l'histogramme.



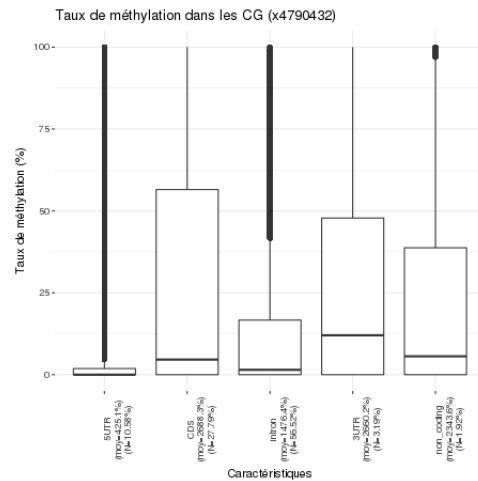
(a) Distribution de la méthylation des CG dans les promoteurs

(b) Distribution de la méthylation dans les gènes

FIGURE 4.2 – Distribution de la méthylation des CG dans les caractéristiques génomiques. Les caractéristiques génomiques représentées sont les promoteurs (a) et les gènes (b). L'axe x représente le taux de méthylation, compris entre 0 et 100% et l'axe y, compte le nombre de CG présents pour la valeur de méthylation. Chaque barre représente un intervalle de méthylation de 5%.

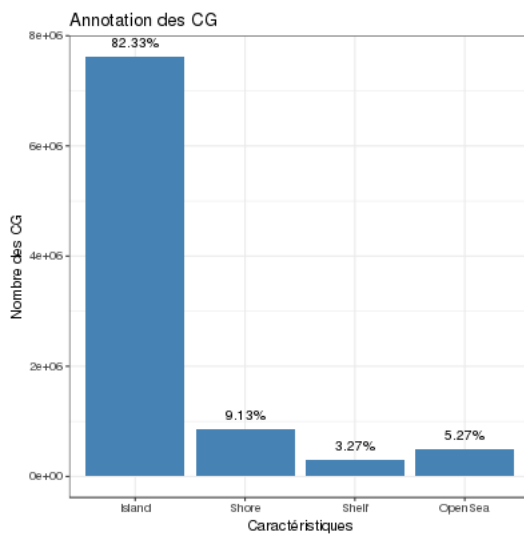


(a) Distribution des CG dans les caractéristiques géniques

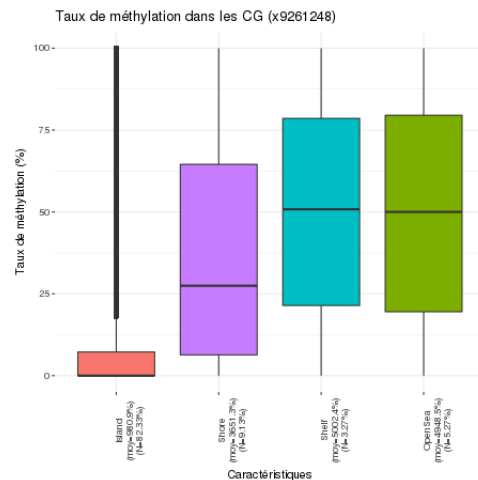


(b) Distribution de la méthylation dans les caractéristiques géniques

FIGURE 4.3 – Répartition des CG et leur taux de méthylation dans les caractéristiques géniques. Le pourcentage de CG présents dans chaque catégorie génique est calculé par rapport à l'ensemble des positions observées dans ces caractéristiques (a). Le taux de méthylation est observé pour chaque caractéristique génique : les 5'UTR, CDS, introns, 3'UTR et régions non codantes (b). La moyenne de méthylation et le nombre de CG sont notés sur l'axe x.



(a) Distribution des CG dans les îlots CpG et les régions flanquantes



(b) Distribution de la méthylation dans les îlots CpG et les régions flanquantes

FIGURE 4.4 – Répartition des CG (a) et leur taux de méthylation (b) dans les îlots CpG et les régions flanquantes. Le nombre de CG présents dans les îlots CpG, shores, shelves et openSea est calculé par rapport à l'ensemble des positions observées dans ces régions. Le taux de méthylation est observé dans chaque région au sein et autour des îlots CpG. Le nombre de CG et la méthylation moyenne et sont notés sur l'axe x.

Nous avons analysé le profil de la méthylation de l'ADN 1kb en amont du gène, au sein du gène et 1kb en aval du gène. Il a été montré qu'au sein du gène, le niveau de méthylation est plus important que dans les régions en 5' et 3' du gène [185]. Notre analyse révèle une diminution globale de la méthylation avant le TSS (*Transcription Start Site*), et une augmentation globale dans le gène. Le profil de méthylation a été observé

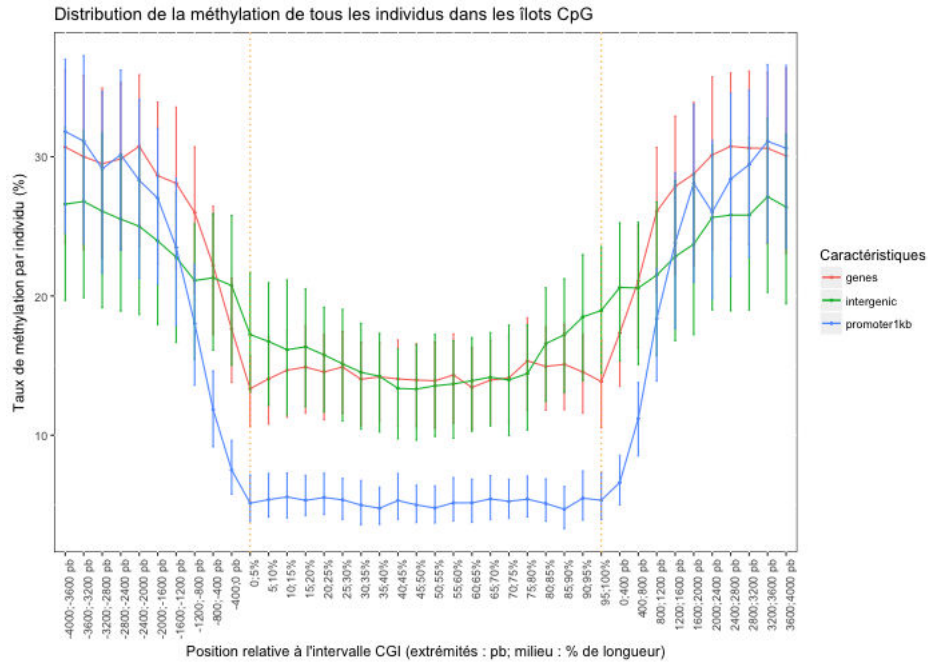


FIGURE 4.5 – Taux de méthylation moyen dans les îlots CpG (CGI) selon les caractéristiques génomiques pour tous les individus. Les points sont le taux de méthylation pour chaque intervalle de position. Les barres d'erreur représentent la valeur minimale et maximale par point. Entre les lignes en pointillé orange, l'intervalle montre les positions des CGI en pourcentage de la taille de l'intervalle CGI. À l'extérieur des lignes en pointillé, les intervalles sont en pb. Au sein des CGI, la méthylation est plus faible dans les promoteurs, tandis qu'elle est similaire dans les régions géniques et intergéniques. Dans les régions flanquantes des CGI, la méthylation augmente lorsqu'on s'éloigne.

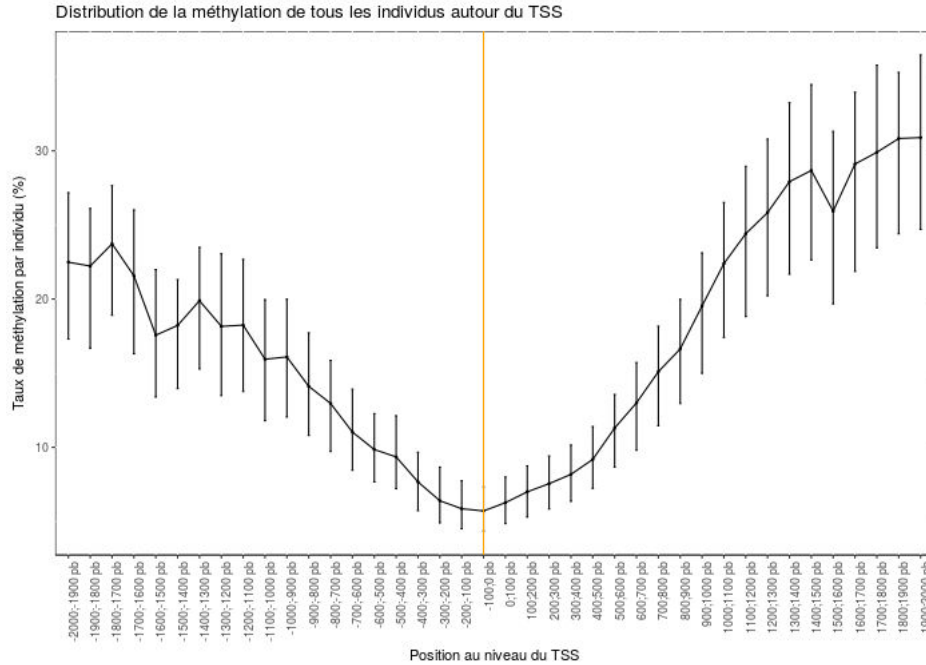


FIGURE 4.6 – Taux de méthylation autour du TSS sur tous les individus RRBS. Chaque taux de méthylation moyen pour tous les individus est représenté par la moyenne (ligne noire). Les barres d'erreur sont les valeurs minimales et maximales au sein des 32 taux de méthylation. La ligne orange montre la moyenne de méthylation dans l'intervalle -100 pb en amont du TSS.

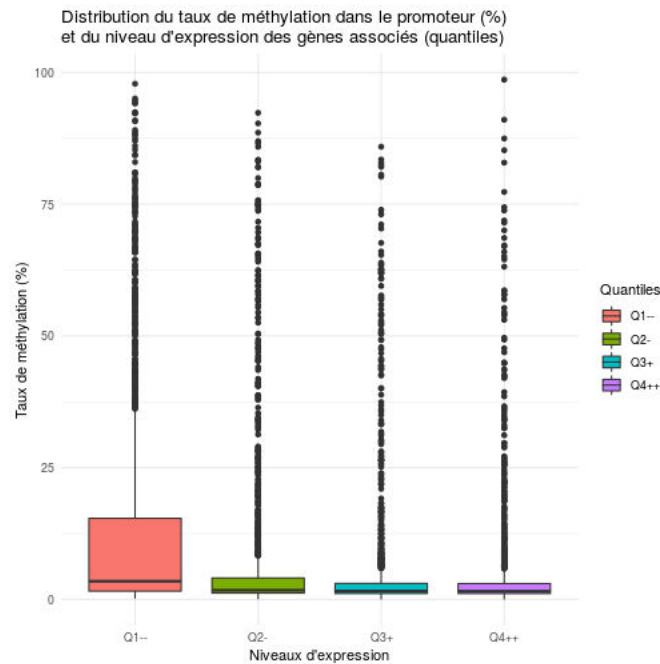


FIGURE 4.7 – Distribution de la méthylation entre quatre quantiles d'expression pour tous les individus. Ces quantiles sont calculés sur les valeurs d'expression de tous les gènes. Le premier quantile d'expression est le moins exprimé et le quatrième quantile est le plus exprimé.

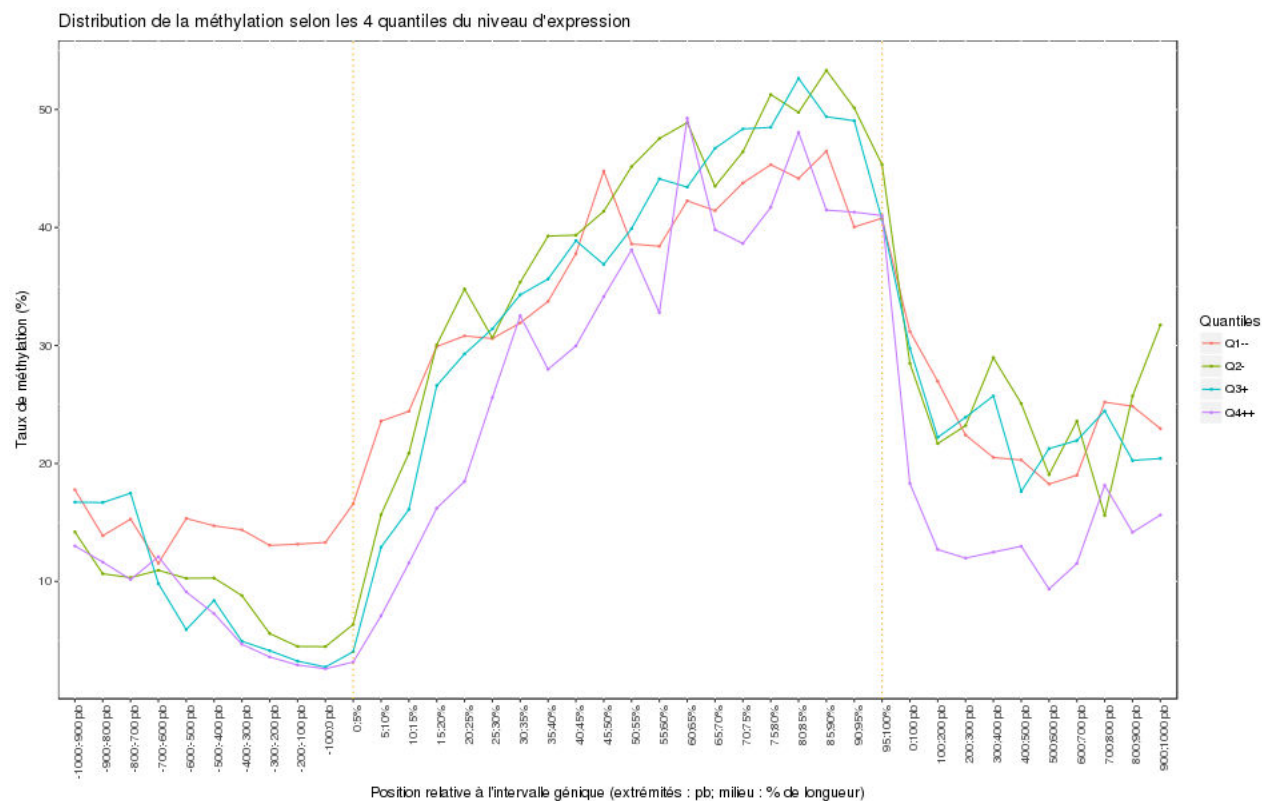


FIGURE 4.8 – Distribution du taux de méthylation selon les niveaux d'expression dans le gène. Les niveaux d'expression sont séparés en 4 quantiles, le quantile 1 (rouge) étant le moins exprimé. Chaque taux de méthylation associé est représenté en amont du gène (extrémité avant la ligne orange), dans le gène (entre les deux lignes oranges) et en aval du gène (extrémité après la ligne orange). À l'extérieur des lignes orange, la distance est donnée en paires de base par rapport à la position du TSS, ou du TTS. Entre les lignes orange, il s'agit d'un pourcentage de la taille de l'intervalle.

pour les quatre quantiles d'expression (Figure 4.8). Le niveau de méthylation dans le quantile le moins exprimé est le plus haut avant le TSS, ce qui confirme une corrélation négative entre ces deux mécanismes. Au sein du gène, la méthylation augmente pour tous les quantiles d'expression, sans montrer de différence nette entre les niveaux d'expression. Néanmoins à la fin du gène, au TTS (*Transcription Termination Site*), le taux de méthylation chute, notamment dans le quantile le plus exprimé, mais reste plus méthylé qu'en amont du gène. Ce profil de méthylation le long du gène est retrouvé sur des données de muscle chez les bovins [218] et sur des données de placenta chez les bovins [209]. La densité de CG détectés grâce au RRBS permet de voir les zones riches en CG, notamment en amont et au début du gène.

4.3 Différentiels de méthylation

Les données nous permettent de réaliser deux types d'analyses : des différentiels de méthylation d'origine génétique, entre les lignées R^+ et R^- , et des différentiels de méthylation d'origine environnementale, entre les conditions contrôle et stress (cf Tableau 2.5 page 79). Les mêmes paramètres que précédemment sont appliqués pour la détection de DMC et de DMR pour chaque condition.

4.3.1 Différentiel de méthylation entre lignées

Les 32 individus sont utilisés pour calculer le différentiel de méthylation entre lignées. En effet, 16 individus sont R^+ et autant sont R^- : 953 DMC sont détectées et 11 DMR (Tableau 4.3). Aucune différence significative n'est observée pour le sens du différentiel de méthylation entre les lignées. En effet, les nombres de positions hyper- ou hypo-méthylées (UP/DOWN) entre les lignées sont équivalents ($p = 0,44$, test binomial exact). À l'origine, les lignées R^+ et R^- proviennent d'une même lignée [5]. Le critère de consommation alimentaire résiduelle sur lequel a été effectuée la sélection divergente n'est pas gouverné par un gène majeur mais par plusieurs gènes. Une partie des régions génomiques qui différencient les lignées portent donc des gènes intervenant dans la variabilité de ce caractère. Ces différences peuvent être de natures diverses, comme la modification d'une protéine affectant sa fonction par exemple. Mais elles peuvent également se traduire par des différences d'expression, avec des gènes actifs dans la lignée R^+ et peu ou pas exprimés dans la lignée R^- , ou inversement. En effet, la lignée R^+ est caractérisée par une thermogénèse induite par l'aliment plus élevée [6] et détourne plus de matières protéiques à des fins énergétiques [20]. Cette lignée est plus économe en aliment à température ambiante mais a des performances limitées à température élevée. On s'attend ainsi à ce que la variation de la méthylation joue un rôle d'activation ou de répression dans des gènes liés à la consommation résiduelle et permette l'adaptation phénotypique de la poule.

Les DMC sont retrouvées dans les îlots CpG pour 456 d'entre elles (47,85%), 218 dans les *shores* (22,88%) et 101 dans les *shelves* (10,60%). De plus, 121 DMC sont associées à 81 promoteurs. Aucun enrichissement n'a été observé, mais la majorité des gènes concernés sont impliqués dans les processus cellulaires et métaboliques (Figure 4.9). On détecte 3 gènes associés à une DMR (gènes JKAMP, MRC2 et ENSGALG00000045029). Le gène JKAMP (*JNK1/MAPK8 Associated Membrane Protein*) est un régulateur de l'activité de MAPK8, qui fait partie de la famille des protéines kinases activées par les mitogènes. Il est impliqué dans la réponse inflammatoire de tissus lors de maladies comme le cancer et dans la réponse au stress [219]. Le deuxième gène annoté ayant une DMR dans le promoteur est MRC2 (*Mannose Receptor C Type 2*) et joue un rôle dans l'environnement extracellulaire. Le gène MRC2 code pour une protéine impliquée dans la croissance osseuse. Il a été montré qu'une délétion de 2 pb dans le gène entraîne notamment des anomalies squelettiques qui résultent d'un remodelage altéré de la matrice extracellulaire au cours de l'ossification chez les bovins [220]. Au sein des lignées, une différence entre la longueur du tarse a été montrée : la lignée la plus faible consommatrice, R^- , a un plus petit tarse que l'autre lignée [183]. Même si l'implication du gène MRC2 n'est pas démontrée dans les différences squelettiques entre les lignées, il est intéressant de noter qu'un gène impliqué dans la croissance osseuse est différenciellement méthylé entre ces deux lignées.

En outre, 480 DMC sont retrouvées au sein de 344 gènes. Il n'y a pas d'enrichissement détecté, et les principaux processus biologiques des gènes sont des processus cellulaires et métaboliques, comme précédemment. Néanmoins, 5 DMR au sein du corps du gène ont un lien avec 7 gènes (L3HYPDH, ENSGALG00000035131, ENSGALG00000032220, MYO18A, POLR2E, TBC1D25 et ARHGAP45). Parmi les 5 annotés, les processus biologiques retrouvés sont divers (dégradation de protéines en acides aminés, migration cellulaire, synthèse

Condition	Différentiel	Nb total	Sens UP	Sens DOWN
R ⁺ /R ⁻	DMC	953	464 (48,69%)	489 (51,31%)
R ⁺ /R ⁻	DMR	11	6 (54,55%)	5 (45,45%)

Tableau 4.3 – Nombre de DMC et de DMR pour le différentiel entre lignées (R⁺/R⁻). Le sens UP signifie que les R⁻ ont des CG plus méthylés que les R⁺ (DOWN : CG plus méthylés chez les R⁺).

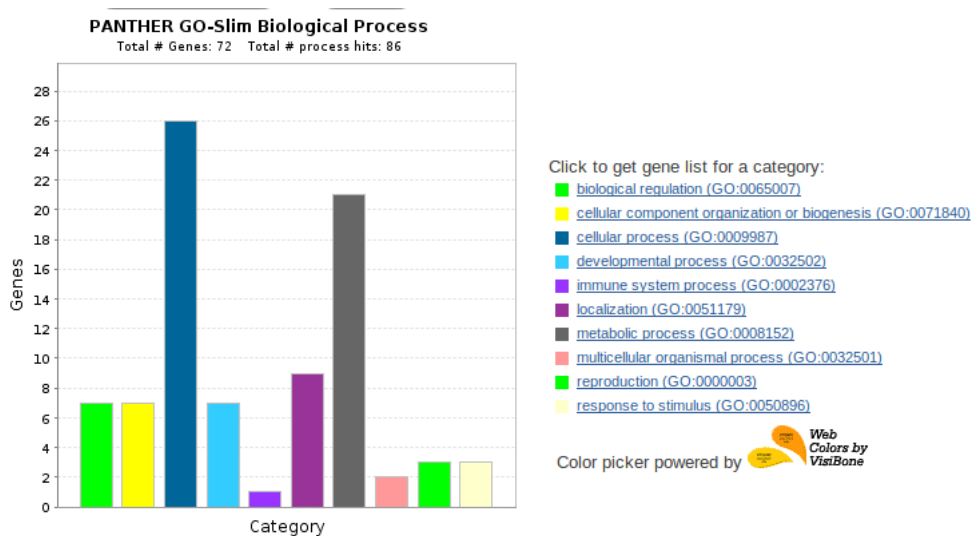


FIGURE 4.9 – Catégories biologiques des DMC entre R⁺ et R⁻ liés à des gènes. Ce graphique représente la répartition des gènes détectés avec une méthylation différentielle dans leur promoteur, et est issu de la base de données PANTHER <http://www.pantherdb.org>.

d'ARN messenger, transport de vésicules à travers la membrane, et réponse immunitaire innée) et ne révèlent pas clairement un lien avec les différences phénotypiques observées entre les lignées.

Les données RNA-seq nous permettant de lier la méthylation avec l'expression ont été analysées par ailleurs (cf partie 4.2.2 page 137), nous n'avons pas encore obtenu les gènes différentiellement exprimés entre les lignées R⁺ et R⁻.

4.3.2 Différentiel entre les contrôles et les stressés

4.3.2.1 Stress à la chaleur

Le différentiel entre le groupe contrôle (CT) et le groupe stressé à la chaleur (HS) est calculé entre 10 individus contre 11 individus, respectivement. Parmi les 801 DMC et 3 DMR détectées, la majorité sont plus méthylées dans la condition HS par rapport au contrôle (dans les deux cas : $p < 2,2 \cdot 10^{-16}$, test binomial exact ; Tableau 4.4). L'analyse montre que 90,76% des DMC et 100% des DMR ont un taux de méthylation plus important en cas de stress chaleur. Ces résultats suggèrent soit que le stress à la chaleur a induit une augmentation globale du niveau de méthylation sur le génome, soit que seules quelques régions spécifiques ont eu une augmentation de leur niveau de méthylation. La distribution des taux de méthylation entre les deux conditions révèle une hyper-méthylation des CG en condition HS (Figure 4.10). En revanche, une étude WGBS chez le porc révèle une hypo-méthylation globale du génome en condition de stress à la chaleur [211]. Cette différence peut être en partie expliquée par les deux technologies utilisées : le WGBS qui cible tout le génome et le RRBS qui cible les régions du génome enrichies en CG. Ce résultat original méritera une analyse plus approfondie quand l'ensemble des données seront disponibles. Nous avons montré précédemment que 18,31% des CG issus de RRBS sont dans les promoteurs (cf Figure 4.1) et sont peu méthylés (cf Figure 4.2), ce qui est cohérent avec la littérature. Or seules 5,46% des DMC détectées sont retrouvées dans les promoteurs, et les résultats montrent un enrichissement des DMC principalement dans les régions intergéniques (+7,24%) (Figure 4.11 a). Le nombre de DMC dans les îlots CpG est également plus faible que le nombre de CG (51,06% ; 82,33% respectivement) (Figure 4.11 b). Ces résultats suggèrent qu'une plus grande proportion de DMC par rapport aux CG sont présentes dans les régions intergéniques, qui incluent notamment les régions répétées qui sont des régions hyper-méthylées [193]. Ainsi, nos résultats révèlent que ces régions condensées sont sensibles à la variation de la méthylation en réponse à un stress thermique. Cependant, des analyses complémentaires seraient nécessaires pour définir dans quelle mesure ces régions répondent à la régulation de l'expression des gènes, et participent à la tolérance à la chaleur.

Les DMC dans les promoteurs sont associées à 48 gènes dont aucun n'est différentiellement exprimé entre les stressés et les contrôles. Cette absence de lien entre une méthylation différentielle dans le promoteur et l'expression du gène associé peut s'expliquer par une faible quantité de gènes différentiellement exprimés ($n = 31$; FDR $\leq 0,05$), et par un nombre de DMC observés dans les promoteurs également limité ($n = 60$; p-valeur ajustée $\leq 0,05$). Les gènes différentiellement exprimés sont issus des analyses d'expression précédentes pour lesquelles un niveau d'expression par condition a été calculé avec le package edgeR [221]. Cependant, deux gènes différentiellement exprimés sont liés à une méthylation différentielle intragénique : PRRX1 et ENSGALG00000043016. Tous les deux présentent un taux de méthylation plus important et un niveau d'expression plus faible dans le groupe stressé. Le seul gène annoté, PRRX1 (*Paired Related Homeobox 1*), est impliqué dans la régulation du devenir des cellules du mésenchyme et dans la genèse d'adipocytes d'après une étude conduite chez les souris [222]. Cette étude suggère que l'augmentation de l'expression du gène PRRX1 dans le tissu adipeux induit une activité des facteurs de croissance et contribue ainsi à une augmentation de l'adipogenèse chez les souris obèses. Il a été montré que l'exposition prolongée à la chaleur des deux lignées entraîne une diminution de poids et un plus faible dépôt de tissu adipeux [6]. Cela s'accorde avec le plus faible niveau d'expression du gène PRRX1 en condition HS. De plus, les DMC associées au corps du gène montrent un enrichissement dans le développement du mésoderme (Figure 4.12). On sait qu'une variation de la méthylation est induite suite à un stress chaleur pendant le développement embryonnaire [126]. Les tissus analysés ici étant des tissus adultes, une analyse approfondie des gènes en question est nécessaire pour interpréter correctement ces différences, subsistant bien après le développement du mésoderme. Les DMR sont toutes plus méthylées dans le groupe HS, mais n'ont pas de lien avec des gènes différentiellement exprimés. Une DMR chevauche la région promotrice et le gène CD93 (*Human cluster of differentiation 93*). Son expression a été analysée dans différents types cellulaires impliqués dans la cascade inflammatoire et dans l'hématopoïèse, renouvellement des cellules sanguines [223][224]. La méthylation de l'ADN a été montrée comme étant associée au devenir

Condition	Différentiel	Nb total	Sens UP	Sens DOWN
CT/HS	DMC	801	74 (9,24%)	727 (90,76%)
CT/HS	DMR	3	0 (0%)	3 (100%)

Tableau 4.4 – Nombre de DMC et de DMR pour le différentiel stressé à la chaleur (HS/CT). Le sens UP signifie que les CT ont des CG plus méthylés que les HS (DOWN : CG plus méthylés chez HS).

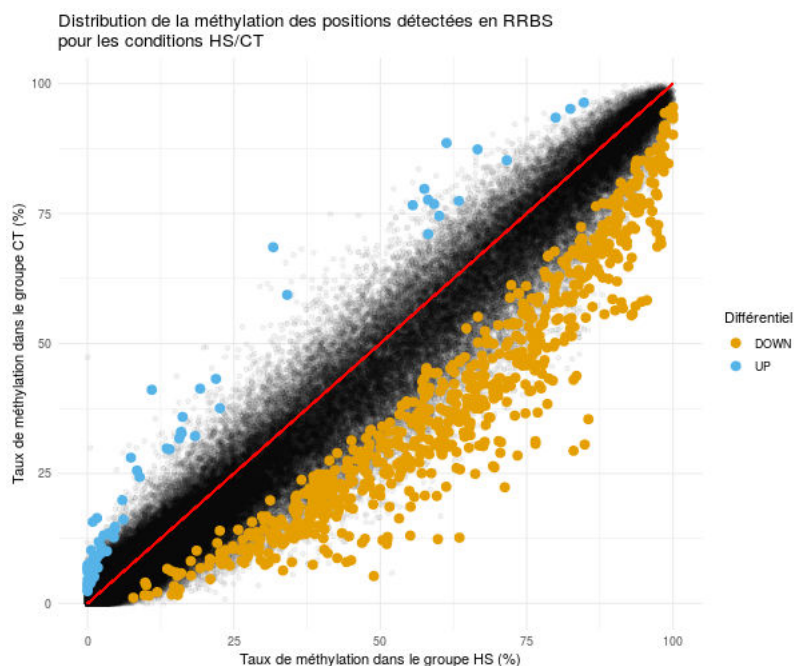
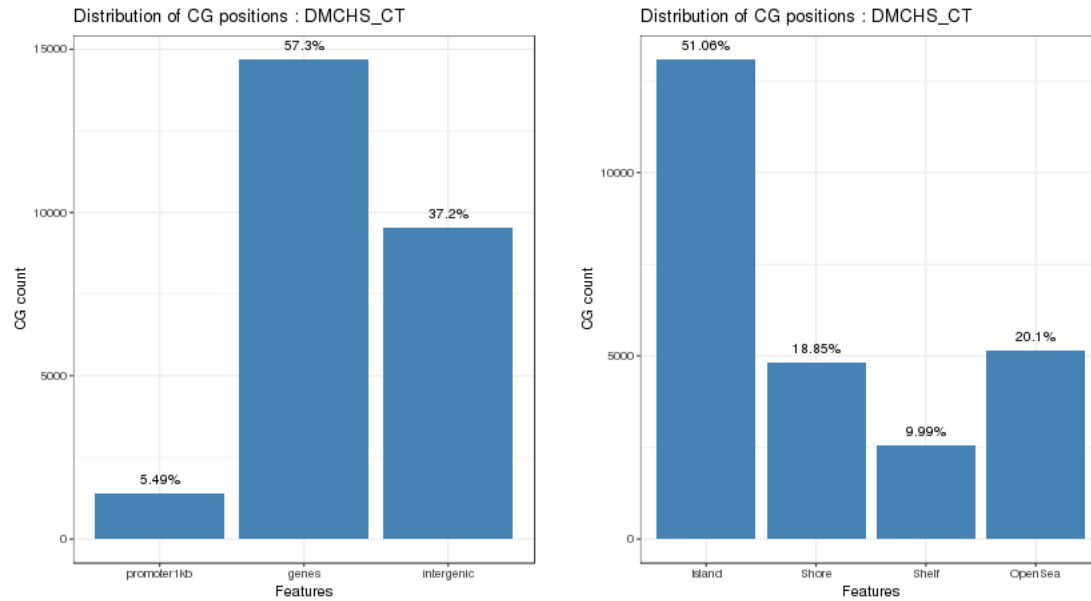


FIGURE 4.10 – Distribution de la méthylation des CG et des DMC pour les conditions CT et HS. Les points de couleur noire représentent la distribution de la méthylation des CG. Les positions qui sont détectées comme différemment méthylées (DMC) sont dessinées en orange lorsque les HS ont des CG plus méthylés que les CT (DOWN), et en bleu, lorsque les CT ont des CG plus méthylés que les HS (UP). La ligne rouge représente les taux de méthylation équivalents dans les deux conditions (elle passe par 50% de méthylation pour les conditions HS et CT).



(a) Distribution des DMC dans les régions génomiques (b) Distribution des DMC dans les îlots CpG et les régions flanquantes

FIGURE 4.11 – Répartition des DMC issues du différentiel CT/HS, dans les régions génomiques (a) et les îlots CpG et les régions voisines (b). Les pourcentages au-dessus de chaque barre d'histogramme représentent la proportion de DMC dans la catégorie fonctionnelle.

	Gallus gallus (REF)	list.genes (▼ Hierarchy NEW! ⓘ)					
PANTHER GO-Slim Biological Process	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
mesoderm development	274	14	4.43	3.16	+	2.25E-04	2.70E-02
↳ developmental process	1392	47	22.50	2.09	+	2.57E-06	6.16E-04
transcription from RNA polymerase II promoter	758	26	12.25	2.12	+	3.86E-04	2.32E-02
Unclassified	9028	114	145.95	.78	-	2.92E-04	2.34E-02

FIGURE 4.12 – Enrichissement des DMC dans le gène pour le différentiel CT/HS. Il est issu de la base de données PANTHER <http://www.pantherdb.org>.

Condition	Différentiel	Nb total	Sens UP	Sens DOWN
CT/FS	DMC	794	29 (3,65%)	765 (96,35%)
CT/FS	DMR	4	0 (0%)	4 (100%)

Tableau 4.5 – Nombre de DMC et de DMR pour le différentiel carencé en énergie (FS/CT). Le sens UP représente les CG plus méthylés chez le groupe CT et le sens DOWN montre les CG plus méthylés au sein du groupe FS.

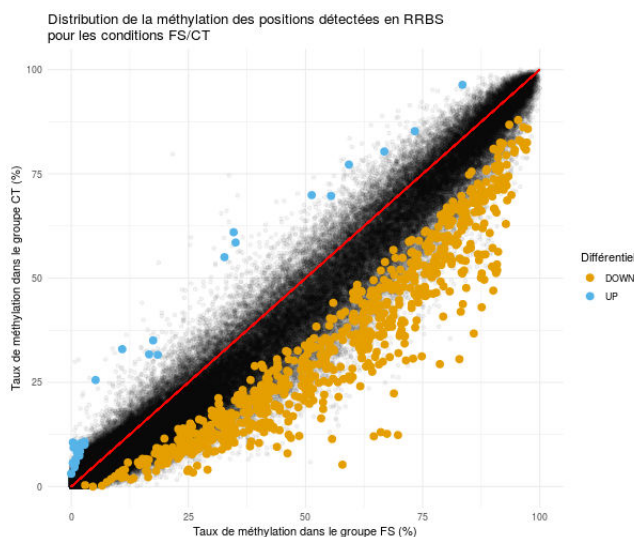
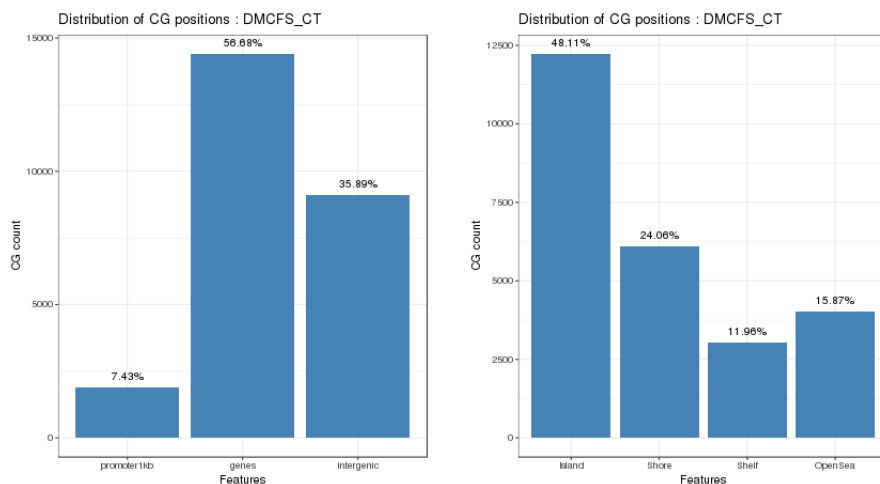


FIGURE 4.13 – Distribution de la méthylation des CG et des DMC pour les conditions CT et FS. Les points de couleur noire représentent la distribution de la méthylation des CG. Les positions qui sont détectées comme différemment méthylées (DMC) sont dessinées en orange lorsque les FS ont des CG plus méthylés que les CT (DOWN), et en bleu, lorsque les CT ont des CG plus méthylés que les FS (UP). La ligne rouge représente les taux de méthylation équivalents dans les deux conditions (elle passe par 50% de méthylation pour les conditions FS et CT).



(a) Distribution des DMC dans les régions génomiques

(b) Distribution des DMC dans les îlots CpG et les régions flanquantes

FIGURE 4.14 – Répartition des DMC issues du différentiel CT/FS, dans les régions génomiques (a) et les îlots CpG et les régions voisines (b). Les pourcentages au-dessus de chaque barre d'histogramme représentent la proportion de DMC dans la catégorie fonctionnelle.

des cellules sanguines dans l'hématopoïèse, et sa perturbation pourrait affecter les populations cellulaires qui constituent le sang périphérique [112][225]. Ces résultats seront à intégrer avec les données issues des analyses de composition sanguine opérées par ailleurs sur les mêmes animaux. Deux autres DMR sont quant à elles incluses au sein de la région génique de leur gène associé. Le gène SMG6 est exprimé dans le cerveau humain et intervient dans la maintenance de l'homéostasie tissulaire et cellulaire lors du développement embryonnaire [226], mais à notre connaissance, aucun lien direct n'a été fait entre ce gène et la réponse au stress, même chez les mammifères. L'ARN long non codant, ENSGALG00000035091, n'est pas annoté.

4.3.2.2 Utilisation d'un aliment pauvre en énergie

Comme pour le dispositif "stress à la chaleur", le différentiel de méthylation en lien avec le stress alimentaire comporte 10 individus contrôle (CT) et 11 individus pour le stress alimentaire (FS). La majorité des DMC et des DMR sont plus méthylées dans la condition du stress par rapport au contrôle (dans les deux cas : $p < 2,2 \cdot 10^{-16}$, test binomial exact ; Tableau 4.5). Les résultats montrent que 96,35% des DMC et 100% des DMR ont une méthylation plus importante en cas de stress alimentaire. Comme précédemment, la distribution des taux de méthylation entre les conditions CT et FS montre une hyper-méthylation des CG (Figure 4.13). Plus de DMC sont retrouvées dans les promoteurs par rapport au différentiel HS/CT (7,43%), mais c'est toujours moins que le nombre de CG détectés au sein des promoteurs (18,31%). Ces DMC CT/FS montrent un enrichissement dans les régions géniques (+4,95%) et des régions intergéniques (+5,93%) (Figure 4.14 a). Comme précédemment, les régions répétées hyperméthylées sont sensibles à la variation de la méthylation en réponse à un stress environnemental. De plus, un nombre plus important de DMC par rapport au CG observés en RRBS est présent dans les régions voisines des îlots CpG, notamment les *shores* (+14,93%) (Figure 4.14 b). La méthylation des CG dans les CGI restant faible, le contraste du niveau de méthylation avec les *shores* hyperméthylés suggère une classe de CGI particuliers. Ce type de CGI est retrouvé dans les gènes les plus exprimés chez l'Homme [214]. Ainsi, nos résultats suggèrent une variation de la réponse du méthylome à un stress alimentaire.

Comme précédemment, les analyses pour détecter les gènes différentiellement exprimés sont issues des données RNA-seq analysées par ailleurs et ont été réalisées avec le package edgeR (cf partie 4.2.2 page 137). Au total, 2700 gènes différentiellement exprimés entre les conditions contrôle et de stress alimentaire ont été détectés (FDR $\leq 0,05$).

Les DMC dans les promoteurs sont associées à 57 gènes dont 7 sont différentiellement exprimés. Même si aucun enrichissement fonctionnel n'est détecté, 5 sur 7 sont plus méthylés et moins exprimés dans le groupe FS : NDEL1, CAMSAP1, ATG9A, DNAJC16, et ADAM11. Ils sont tous impliqués dans les processus neuronaux. Cependant, le gène DNAJC16 (*DNAJ Heat Shock Protein Family (Hsp40) Member C16*), fait partie des gènes induisant des protéines de choc thermique (Hsp, *Heat shock protein*). Confrontées à un changement environnemental, par exemple à une température brusquement élevée, la plupart des cellules activent l'expression des protéines de choc thermique, sous le contrôle des facteurs de transcription de choc thermique (HSF, *Heat Shock Factor*) [53]. Ce sont des protéines chaperonnes qui s'associent avec d'autres protéines ou ARN, et empêchent l'accumulation de protéines anormales en aidant à conformer correctement les polypeptides ou en les dirigeant vers le protéosome qui les détruit [227]. Une étude a révélé que le gène DNAJC16 était différentiellement méthylé en fonction de la consommation de thé chez les femmes [228]. Le lien entre la méthylation de l'ADN de ce gène et la consommation de thé chez les femmes a permis de révéler un changement dans le métabolisme des œstrogènes, et aurait un impact en diminuant l'inflammation [228]. Ce gène a été également observé comme étant plus exprimé dans les cellules cancéreuses du foie chez l'Homme lors d'un traitement au thé vert [229].

Les 2 autres DMC dans le promoteur liées à un gène plus exprimé sont plus méthylées : ENSGALG0000006899 et S100A16. Ce dernier stimule la production des adipocytes et inhibe l'ostéogenèse [230]. Au sein des deux lignées, une étude a montré que l'utilisation d'un aliment moins énergétique abaisse le poids corporel [20]. Cependant, Chen et al. ont montré que l'hyperméthylation des CG du gène S100A16 est la règle dans les tissus normaux où S100A16 est exprimé, et qu'une corrélation négative avec l'expression est impliquée dans le cancer des poumons chez l'Homme [231]. Dans un développement normal, la méthylation de l'ADN ne semble pas être corrélée à l'expression de ce gène, ce qui suggère que d'autres facteurs, épigénétiques ou non, peuvent intervenir dans la régulation de l'expression de ce gène.

Un grand nombre de DMC liées à des gènes différentiellement exprimés sont détectées dans le corps du gène ($n = 62$), mais aucun enrichissement n'est trouvé. La majorité de ces gènes sont plus méthylés et plus exprimés dans le groupe FS ($n = 41$), et un gène est moins exprimé et moins méthylé dans le même groupe. Le reste des gènes a une corrélation négative entre la méthylation et l'expression (18 sont plus méthylés et moins exprimés dans le groupe FS, et 2 sont moins méthylés et plus exprimés dans le même groupe). Cela suggère soit une régulation de l'expression via d'autres mécanismes, épigénétiques ou non, soit une relation complexe entre méthylation au sein du gène et régulation de l'expression, comme nous l'avons vu précédemment. Il serait intéressant d'analyser de manière plus approfondie ces liens expression / méthylation au sein des gènes, notamment en discernant les dinucléotides CG présents dans les premiers intron ou exon, et dans les introns et exons suivants. Cependant, le gène BCAS3 (*Breast Carcinoma Amplified Sequence 3*), différentiellement exprimé, comporte une DMR dans son corps de gène. Les individus stressés ont un taux de méthylation plus important et un niveau d'expression faible au sein du gène. Une étude a identifié qu'un site CG associé avec le gène BCAS3 présente une méthylation différentielle significative chez les porteurs de l'allèle à risque d'obésité FTO [232]. Le taux de méthylation augmente chez les individus obèses, suggérant que l'effet de l'allèle à risque peut, au moins en partie, être influencé à travers des altérations épigénétiques. Néanmoins, l'étude n'a pas permis de révéler un lien clairement établi entre le CG méthylé et l'obésité ou l'allèle FTO. Le gène BCAS3 est un co-activateur d'œstrogènes induisant une sur-expression dans le cancer du sein [233].

4.3.2.3 Conclusion

Parmi les DMC obtenues en analysant les stress à la chaleur et alimentaire, 106 sont communes : 3 où le niveau de méthylation est plus important dans le cas contrôle et 103 pour lesquelles la méthylation est plus grande dans le groupe stressé. Les DMC communes peuvent être impliquées dans la réponse au stress, sans être spécifiques à une réponse à une déficience énergétique ou une tolérance à la chaleur. Cependant, aucun lien différentiel de méthylation dans le promoteur et le corps du gène et d'expression n'est commun dans les deux conditions. Seuls 5 gènes différentiellement exprimés sont communs aux deux conditions de stress : HBBA, HBM, HBA1, H1F0 et ENSGALG00000044985 (ARN long non codant). Les premiers gènes (HBBA, HBM, HBA1) codent pour des hémoglobines, qui sont des protéines majeures présentes dans les érythrocytes et transportent l'oxygène entre tous les tissus. Une étude récente montre que l'hémoglobine neuronale affecte la réponse des cellules dopaminergiques au stress chez l'Homme [234]. La sur-expression de l'hémoglobine augmente la sensibilité des neurones à un stress neurochimique, qui est induit, notamment par une diminution du niveau de H3K9me2 et de H3K9me3. H3K9me3 est liée à la répression de la transcription et inhibe les gènes qui contrôlent la réponse cellulaire aux dommages causés à l'ADN et à différents types de stress [235]. Ainsi, une régulation épigénétique induit une variation de l'expression des gènes, en réponse à un stress. De plus, le gène H1F0, détecté dans notre dispositif EpiBird (cf partie 3.1 page 83), est également connu pour être exprimé dans les érythrocytes polychromatiques chez le poulet [236]. C'est un gène qui code pour l'histone de liaison H5 (variant de H1) de façon indépendante de la réplication et dont l'expression est élevée au sein des globules rouges du poulet [237]. Ces histones de liaison sont nécessaires à la condensation du nucléosome, ce qui pourrait suggérer un changement dans la conformation de la chromatine et une expression génique modifiée. Cependant, aucune étude n'a montré de lien entre l'expression de H1F0 et la réponse à un stress environnemental.

Dans les deux cas de stress environnementaux, la majorité de DMC sont présentes dans les régions riches en CG et dans les promoteurs. Elles montrent une méthylation plus importante dans les conditions de stress, sur plusieurs gènes, impliqués dans plusieurs types de processus biologiques. Même si beaucoup de gènes détectés sont liés au fonctionnement de l'hypothalamus, quelques uns interviennent dans la régulation du métabolisme lipidique (DNAJC16, S100A16, BCAS3), ce qui suggère une régulation de l'expression pour l'adaptation des animaux à la carence en énergie. De plus, l'absence de lien différentiel entre la méthylation et l'expression peut être expliquée par la faible détection de gènes enrichis en processus impliqués dans la réponse aux différents stress environnementaux.

Les résultats décrits précédemment ont été obtenus en fin de thèse, et ils nécessitent une analyse plus approfondie. En particulier, les analyses issues des autres tissus (foie et sang) viendront compléter la caractérisation du méthylome en réponse aux stress et permettront d'aller plus loin dans la compréhension des mécanismes en jeu. Toutefois, un premier résultat original et sans doute important se révèle déjà et méritera qu'on s'y attarde : l'augmentation globale du niveau de méthylation des CG, observée dans les deux types de stress.

Chapitre 5

Synthèse, conclusion générale, perspectives

5.1 Discussion sur les données et les méthodes utilisées

Au cours de la thèse, deux projets de recherche ont généré deux types de données de méthylation : des données tout-génome et des données d'une sous-représentation du génome, chez la poule. Même s'il ne s'agit pas des mêmes dispositifs expérimentaux, il nous a semblé intéressant de comparer les résultats "bruts" de ces données de séquençage, d'analyser la répartition des CG sur le génome, et de réaliser une sous-sélection des positions communes aux deux types de données pour en comparer les résultats. La première partie de cette synthèse discute de la génération des données de séquençage et des limites inhérentes aux 2 techniques de séquençage. De plus, au cours de la thèse, j'ai eu l'opportunité de réaliser une mobilité au sein du Centre National d'Analyses Génomiques (CNAG-CRG) situé à Barcelone, en Espagne. Ce stage s'est axé sur l'utilisation du pipeline de méthylation développé dans l'équipe EpiTeam, qui m'accueillait, pour l'analyse de nos données WGBS de poule, puis sur la comparaison avec les résultats obtenus à partir du pipeline développé dans le cadre de la thèse. Ainsi, la seconde partie de la synthèse discute des méthodes utilisées et choix opérés qui permettent de valider les étapes d'analyses mises en place pour le pipeline développé pendant la thèse.

5.1.1 Données de séquençage

Dans le but de caractériser la méthylation de l'ADN le long du génome, plusieurs techniques sont utilisées (voir revue [238]), notamment les techniques de séquençage à haut-débit qui permettent de générer de plus en plus de données, en moins de temps (cf partie Introduction 1.3 page 61). Le séquençage après conversion au bisulfite est la méthode la plus courante pour distinguer les C méthylés des non méthylés. Comme souligné précédemment, le RRBS ne permet d'accéder qu'à une sous-représentation des CG du génome, mais à un coût bien moindre que le WGBS. Pour un même investissement de départ, il est possible d'obtenir des résultats sur un plus grand nombre d'individus. Au cours de ma thèse j'ai analysé des données des deux types : RRBS et WGBS. Bien que les questions scientifiques soient différentes et que les échantillons soient distincts (lignées, âges des individus, tissus analysés), il m'a semblé intéressant de comparer les deux types de données, en termes de couverture et de profondeur, mais également de représentation des régions fonctionnelles analysées. Dans cet objectif, j'ai comparé la répartition des CG détectés dans les deux jeux de données en prenant en compte tous les CG détectés au sein des échantillons. Puis les résultats obtenus dans les deux types d'analyse ont été comparés sur un sous-ensemble de positions RRBS détectées également en WGBS.

L'utilisation du séquençage après conversion au bisulfite comporte de nombreux défis. La conversion réduit la complexité du génome à 3 nucléotides (à l'exception des dérivés de la 5mC), et ainsi l'alignement des lectures peut devenir plus complexe. De plus, la conversion au bisulfite mène à la fragmentation de l'ADN, ce qui rend l'amplification de longs fragments difficile et peut potentiellement générer des erreurs de lectures. Lors de la préparation de la librairie, la quantité d'ADN requise est d'environ quelques dizaines de nanogrammes pour le WGBS et quelques centaines pour le RRBS [238]. Sur nos données, 1 μg d'ADN en WGBS et 1 μg d'ADN ont été utilisés pour la préparation des librairies pour chaque échantillon.

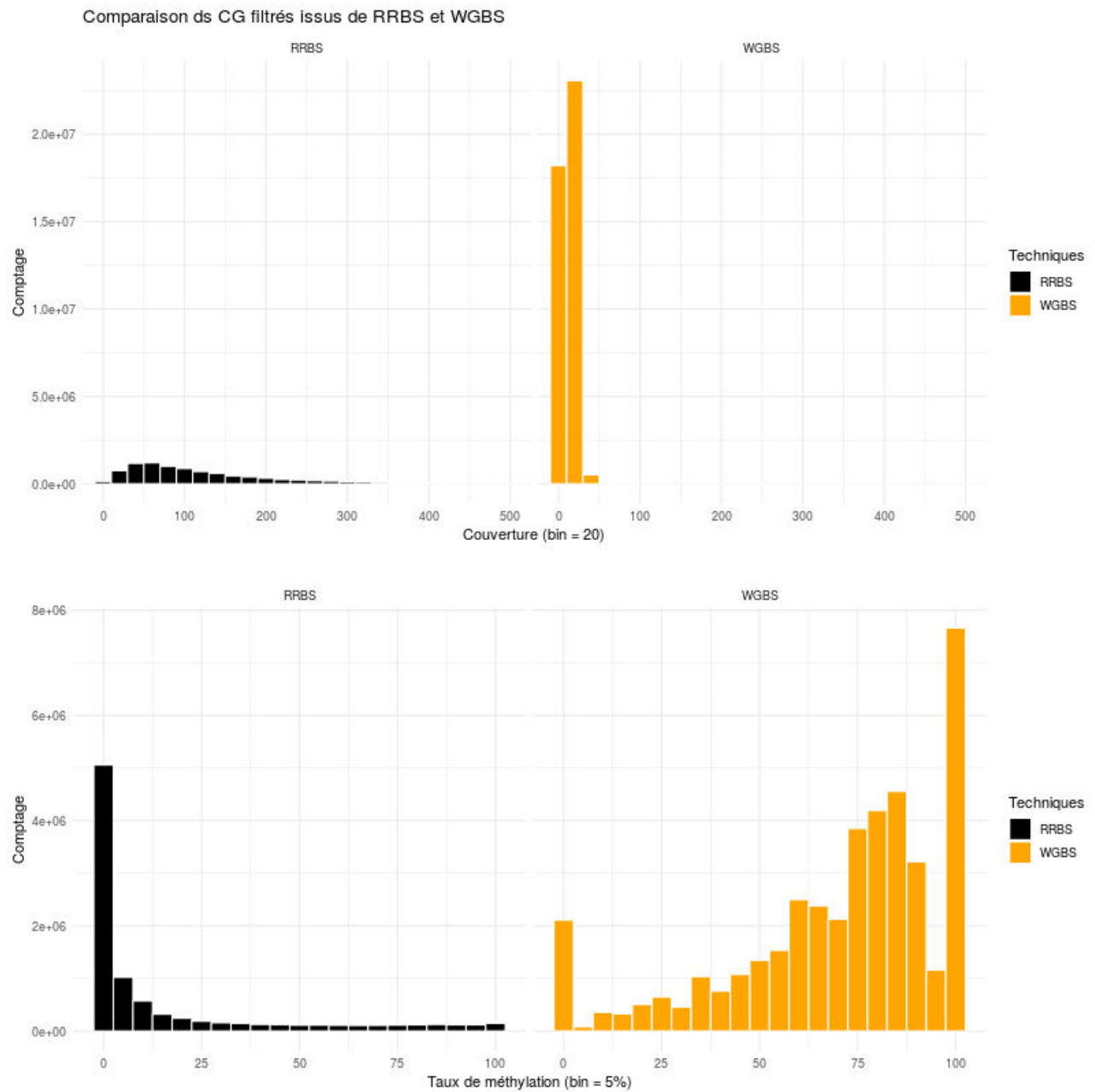


FIGURE 5.1 – Comparaison des positions de CG filtrés issues de RRBS et WGBS. Les deux histogrammes du haut révèlent la distribution de la couverture pour tous les CG et les deux histogrammes du bas montrent la répartition du taux de méthylation. En orange sont représentés les CG issus du séquençage WGBS et en noir, ceux du séquençage RRBS.

En WGBS, nous avons obtenu une couverture, et donc un nombre de CG analysables, compatible avec ce qui était attendu avec la profondeur de séquençage réalisée. Cependant, en RRBS, le nombre de CG détectés est inférieur à ce qui était attendu. Après digestion enzymatique, la succession des purifications de l'ADN pendant la préparation de la librairie fait perdre beaucoup de matériel génétique. De plus, la taille des adaptateurs est de 120 pb, ce qui est très proche de la taille des fragments d'ADN à sélectionner (40-100 pb). La fraction à récupérer correspond donc à la fenêtre de taille 160-220 pb, en comptant les 120 pb d'adaptateurs liés aux extrémités des fragments d'ADN génomique digéré. Cela rend la sélection des fragments plus difficile et peut entraîner des contaminations par des dimères d'amorce. Des purifications supplémentaires ont donc dû être réalisées, diminuant encore la quantité de matériel disponible pour le séquençage. En tenant compte des contraintes financières de ce programme, nous devions regrouper un plus grand nombre d'individus par ligne de séquençage que la plupart des analyses classiques de RRBS. Nous avons donc souhaité optimiser au maximum le nombre de CG détectés, en sélectionnant une fraction de très petits fragments (cf partie 2.2.1 page 75). Néanmoins, cela a été réalisé au détriment de la facilité de récupération de la fraction à séquencer. Il s'agissait de la première expérience de RRBS réalisée sur la plate-forme choisie, en collaboration avec le laboratoire, et le protocole est en train d'être amélioré pour parvenir à des résultats plus satisfaisants en routine.

Les données sont comparées à partir des 9261248 CG filtrés sur les 32 individus RRBS (32*289414) et à partir des 41880699 CG filtrés sur les 7 embryons WGBS (7*5982957). La seule différence entre les analyses qui ont servi à détecter le taux de méthylation de ces CG est l'application d'un minimum de couverture différent : 12X pour le RRBS et 5X pour le WGBS. Ce minimum de couverture est choisi en fonction des données obtenues et des recommandations préconisées dans des études antérieures [208][192]. Nous avons fait le choix de comparer ces données après les filtres en dépit de ce minimum de couverture différent car ce sont ces données filtrées qui sont utilisées pour les analyses décrites dans les chapitres 3 et 4 (Figure 5.1). En moyenne, la couverture des positions RRBS est environ 10 fois plus grande que celle du WGBS (135X et 12,2X respectivement). Après les filtres, la majorité des CG détectés ont un taux de méthylation moyen de 13,83% pour le RRBS tandis que la méthylation des CG détectés par le WGBS varie plus avec un taux de méthylation moyen proche de 71,79%. Cela s'explique par le séquençage tout génome qui ne cible pas de régions particulières dans le génome et donc contient des CG méthylés pour la plupart, dans les régions hors îlots CpG. Cette variation plus grande de la méthylation dans les échantillons embryonnaires peut aussi s'expliquer par l'hétérogénéité de l'échantillon. En effet, les embryons analysés sont un amas de cellules provenant de tous les tissus. Une population mixte de cellules avec plusieurs états de méthylation peut avoir un effet de dilution de la méthylation et donc affecter l'homogénéité des niveaux de méthylation détectés [239].

Il est également intéressant d'observer la répartition des CG dans les fractions enrichies en CG du génome. Il faut noter que les pourcentages d'îlots CpG présents dans les promoteurs sont similaires entre le génome humain et celui de la poule (12,44% et 16,44% respectivement) (calcul réalisé à partir des génomes humain, GCA_000001405.27) et poulet (GCA_000002315.3). Dans les régions riches en CG (îlots CpG, *shores* et *shelves*), le nombre de CG détectés en RRBS décroît au fur et à mesure que l'on s'éloigne des îlots CpG, tandis qu'en WGBS, leur nombre augmente (Figure 5.2). Environ 5% des CG détectés par WGBS sont présents dans des îlots CpG tandis qu'en RRBS, 82% des CG y sont annotés. Cela vient du fait que l'enzyme MspI cible les sites CCGG et permet d'obtenir des fragments riches en CG. Ceci valide les résultats observés, avec un taux de méthylation nettement moins important en RRBS qu'en WGBS à cause de la nature des régions observées. Cependant, le nombre de CG observés est moins important en RRBS qu'en WGBS. La majorité des CG du génome sont dans l'*openSea*, les îlots CpG ne représentant que 2,78% du génome (calcul réalisé sur la taille de tous les îlots CpG du génome divisée par la taille du génome de référence galgal5).

Afin de déterminer si on observe le même profil de répartition des CG dans les deux techniques, un sous-ensemble de positions communes est analysé. Sur l'ensemble des CG analysés dans cette thèse, issus du RRBS ou du WGBS, environ 95% sont uniquement observés en WGBS, 4% uniquement en RRBS et 0,4% dans les deux techniques (Figure 5.3). Ces derniers nous permettent d'établir un profil de méthylation de type RRBS à partir de "vraies" positions WGBS (Figure 5.4). De façon similaire à la figure 5.1, une majorité de CG détectés ont un taux de méthylation compris entre 0 et 5%. Même si ce ne sont pas les mêmes données issues du même dispositif, il est possible de "générer des données RRBS", sur une sous-représentation du génome, à partir de données WGBS. Cela permet d'obtenir un jeu de données virtuel à partir des données réelles pour analyser les régions du génome enrichies en CG, et les plus étudiées.

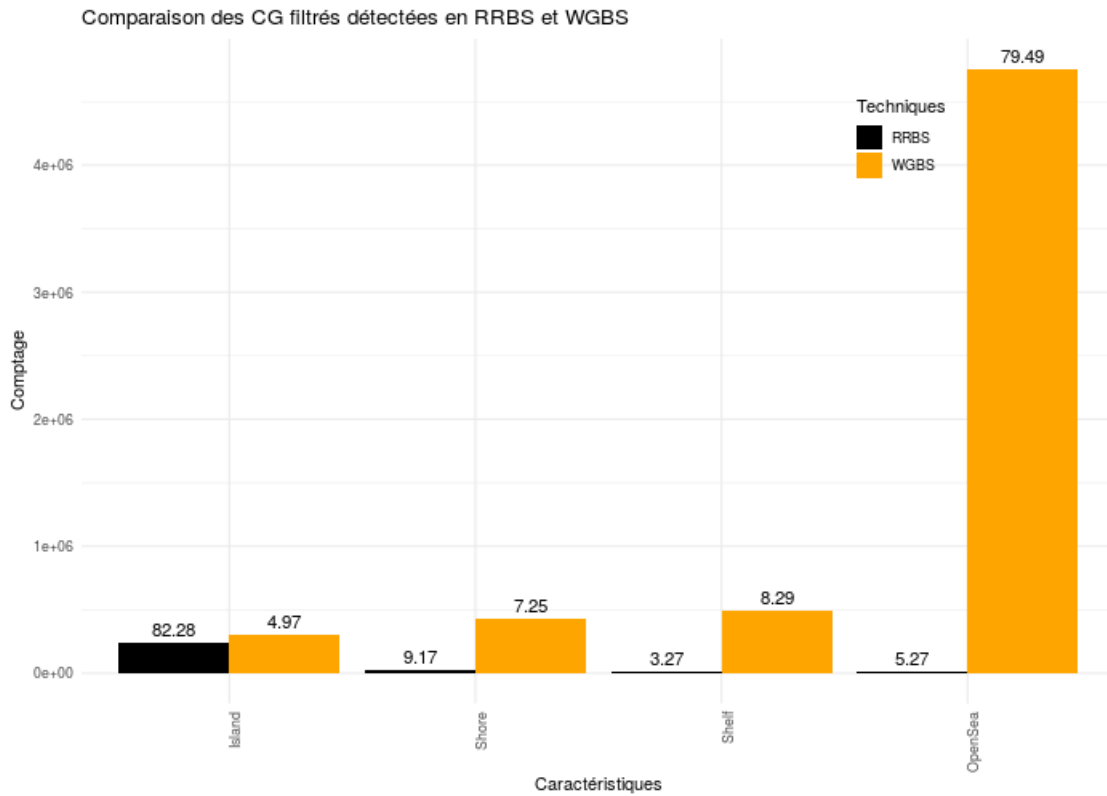


FIGURE 5.2 – Comparaison des positions détectées et annotées. Le nombre de CG annotés dans les CGI, *shores*, *shelves* et *openSea* est soit issu de la technologie WGBS (en orange), soit du RRBS (en noir). Le chiffre au-dessus de chaque barre d'histogramme est le pourcentage de CG dans une région parmi tous les CG détectés au sein d'une technique.



FIGURE 5.3 – Distribution des CG issus des deux techniques de séquençage. En orange, les CG filtrés provenant du WGBS et en noir, ceux du RRBS.

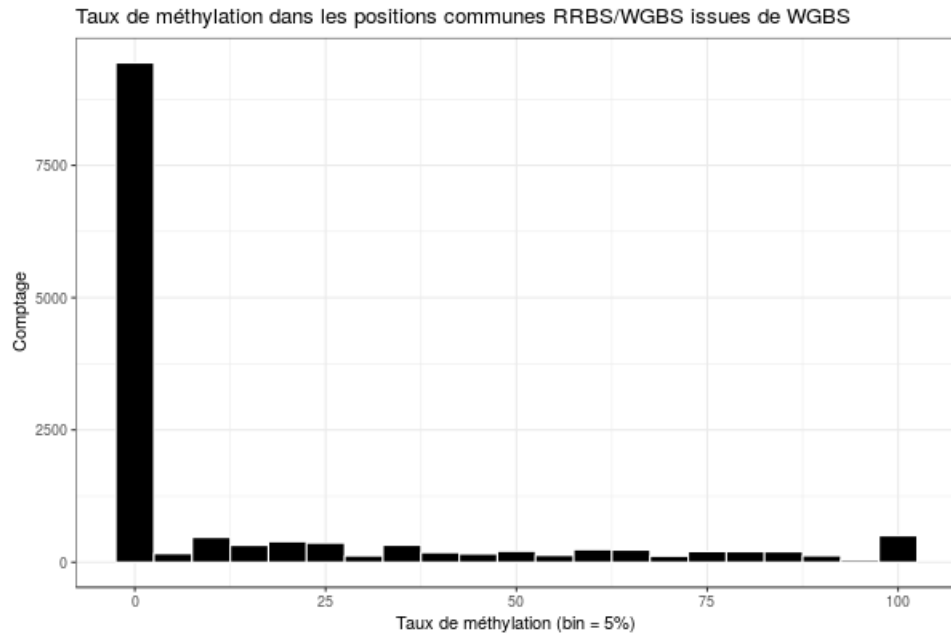


FIGURE 5.4 – Distribution de la méthylation des CG communs RRBS/WGBS à partir des CG WGBS. L'axe x montre le taux de méthylation des CG, pour des intervalles de 5%, et l'axe y compte le nombre de CG représentés.

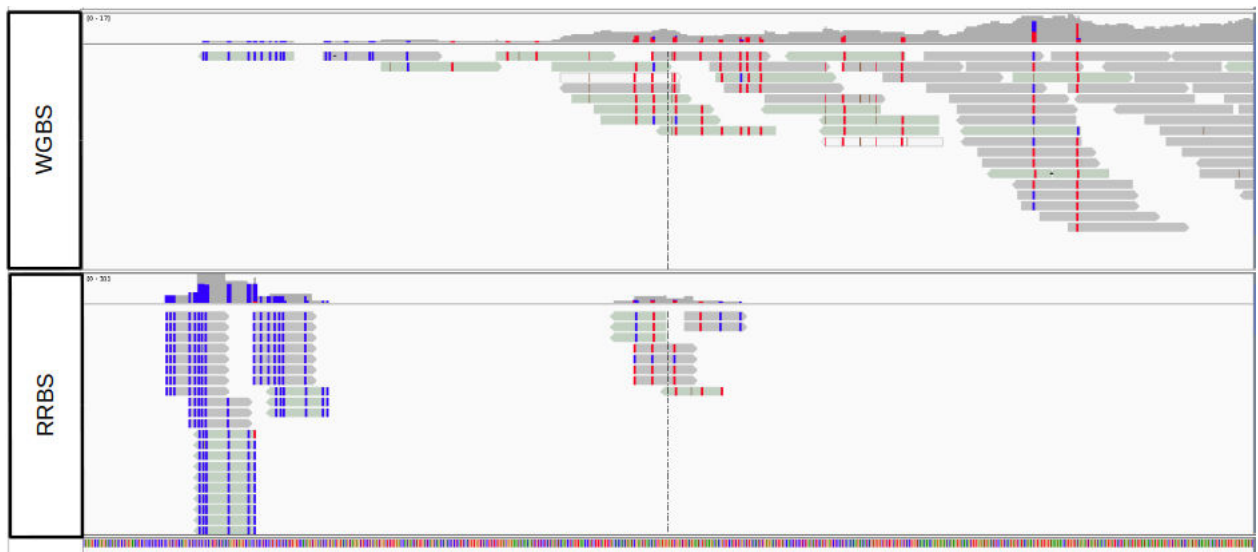


FIGURE 5.5 – Visualisation d'une région génomique analysée en RRBS et WGBS. Un embryon séquencé en WGBS est représenté en haut et un individu séquencé en WGBS, sur la même région, est présenté en bas de la figure.

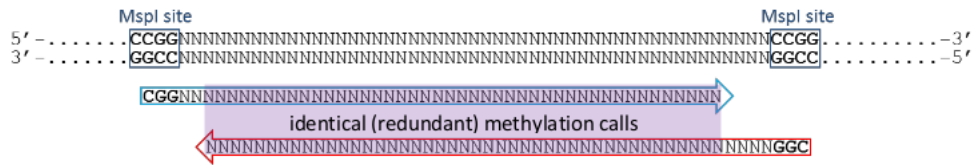


FIGURE 5.6 – Redondance de l’information de méthylation sur des lectures *paired-end* (figure issue du Guide RRBS [155]).

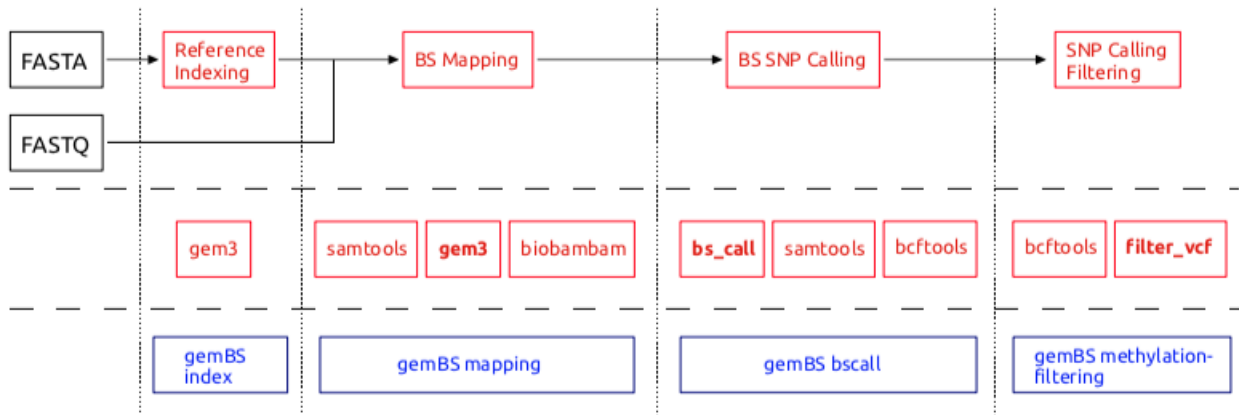


FIGURE 5.7 – Étapes du pipeline gemBS. La première ligne en rouge décrit et titre les étapes des analyses des données. La seconde ligne en rouge note les logiciels au cœur des étapes du pipeline. La dernière ligne en bleu reprend le pipeline gemBS.

Une différence entre deux échantillons provenant de données RRBS et WGBS est observée (Figure 5.5). Sur la même région du génome, le nombre de CG couverts est plus important en WGBS. Dans les données RRBS, on aperçoit beaucoup de lectures qui s'alignent à un endroit du génome tandis que ces mêmes positions ne sont presque pas couvertes par le WGBS. Néanmoins, le deuxième groupe de lectures, au niveau de la ligne verticale en pointillés, est couvert de façon similaire entre le RRBS et le WGBS. Comme nous l'avons vu précédemment, la couverture en RRBS est en théorie supérieure à celle du WGBS, et la région commune aurait dû avoir une profondeur de séquençage supérieure en RRBS. Un des problèmes rencontrés en RRBS est que le processus de sélection de taille des fragments n'est pas aussi efficace dans la réalité qu' *in silico*. De plus, à cause de la sélection de fragments de taille très courte, puisque les fragments étaient ciblés dès 40 pb et que la sélection n'exclut pas à 100% les fragments de part et d'autre de la fenêtre visée, nous avons récupéré un nombre non négligeable de fragments dont la taille est inférieure à 40 pb. Si la longueur de la lecture est plus longue que le fragment digéré par MspI, le séquenceur peut continuer à lire dans la séquence de l'adaptateur en 3'. Une fois les séquences d'adaptateurs éliminées des lectures lors de la première étape du pipeline, la séquence d'ADN restante n'est pas prise en compte dans le processus d'alignement si elle est trop courte (minimum 20 pb). Cela peut expliquer en partie le plus faible nombre de lectures qui s'alignent en RRBS (cf revue [155]). Mais la taille des lectures et leur nature (*single-end*) est la raison principale de cette baisse d'efficacité d'alignement pour le RRBS : lors du séquençage, le fragment d'ADN peut être séquençé soit à partir de ses deux extrémités (*paired-end*), soit à partir d'une seule (*single-end*).

Sur nos données, le WGBS a été séquençé en *paired-end* et le RRBS en *single-end*. Une lecture *paired-end* provient de l'amplification soit du brin sens (OT), soit du brin anti-sens (OB) de l'ADN. Les autres extrémités séquencées sont des séquences des brins complémentaires, CTOT ou CTOB. Ainsi, si les fragments séquencés ont une taille inférieure à 2 fois la longueur de lecture, ce qui est le cas en RRBS, les lectures *paired-end* peuvent se chevaucher et une information de méthylation redondante pour le même brin est détectée (Figure 5.6). En raison de cette information redondante sur la méthylation, qui est supprimée par Bismark, séquencer des fragments courts en *paired-end* ne double pas la quantité d'information de méthylation recueillie. Le séquençage *single-end* avec le même nombre de lectures qu'en *paired-end* est plus susceptible de fournir un plus grand nombre d'informations de méthylation, à condition que la longueur de la lecture soit suffisante pour un bon alignement. Ainsi, un séquençage en *single-end* pour du RRBS est recommandé pour éviter ce biais. En WGBS, le *paired-end* permet d'augmenter le taux d'alignement. Sur nos données, le taux moyen d'alignement est de 70,37% pour les embryons issus du WGBS et de 51,62% pour les individus issus du RRBS. Cette diminution du taux d'alignement en RRBS peut s'expliquer par la taille des fragments plus petits, comme décrit précédemment, mais également par la contamination des séquences d'amorces. Dans le pipeline, l'option *paired-end* ou *single-end* est présente. L'outil Trim Galore supprime les séquences d'adaptateurs en 3' pour le *single-end* et en plus en 5' sur la deuxième lecture de la paire en *paired-end* (cf partie Introduction 1.3.3.1 page 65). Enfin, alors que le taux d'alignement issu de données provenant de DNA-seq est proche de 100% (cf revue [240]), le taux d'alignement en WGBS est plus faible (70%). Il peut s'expliquer par la dégradation de l'ADN due au bisulfite de sodium et par la diminution de la complexité du génome (diminution de la spécificité de l'alignement à cause de la conversion de la plupart des C en T par le bisulfite de sodium).

5.1.2 Pipeline d'analyses

Dans le cadre de ma mobilité au CNAG-CRG de Barcelone, j'ai pu utiliser un pipeline dédié à la détection de la méthylation en WGBS testé sur des données obtenues à partir de cellules humaines de la moelle épinière [241]. Un des objectifs était de tester ce pipeline, initialement développé pour le génome humain, sur nos données aviaires. Une étape supplémentaire de ce pipeline par rapport au pipeline développé pendant la thèse est la détection de SNP à partir de lectures traitées au bisulfite. Dans notre étude sur les embryons, grâce à nos données issues de DNA-seq, nous avons pu identifier les SNP au préalable pour chaque embryon. Cela nous a permis d'établir une liste de SNP fiables, qui ont ensuite pu être supprimés de l'analyse (cf partie 1.3.3.1 page 67). Néanmoins, obtenir les séquences génomiques lors de programmes utilisant des données BS-seq n'est pas toujours possible. C'est pourquoi le pipeline gemBS développé au CNAG identifie les SNP à partir des lectures issues de BS-seq. De plus, le pipeline gemBS diffère du nôtre par l'association d'un aligneur à haute performance appelé GEM3 [242] et d'un logiciel de détection de variants appelé BScall. Les étapes du pipeline sont décrites dans la figure 5.7.

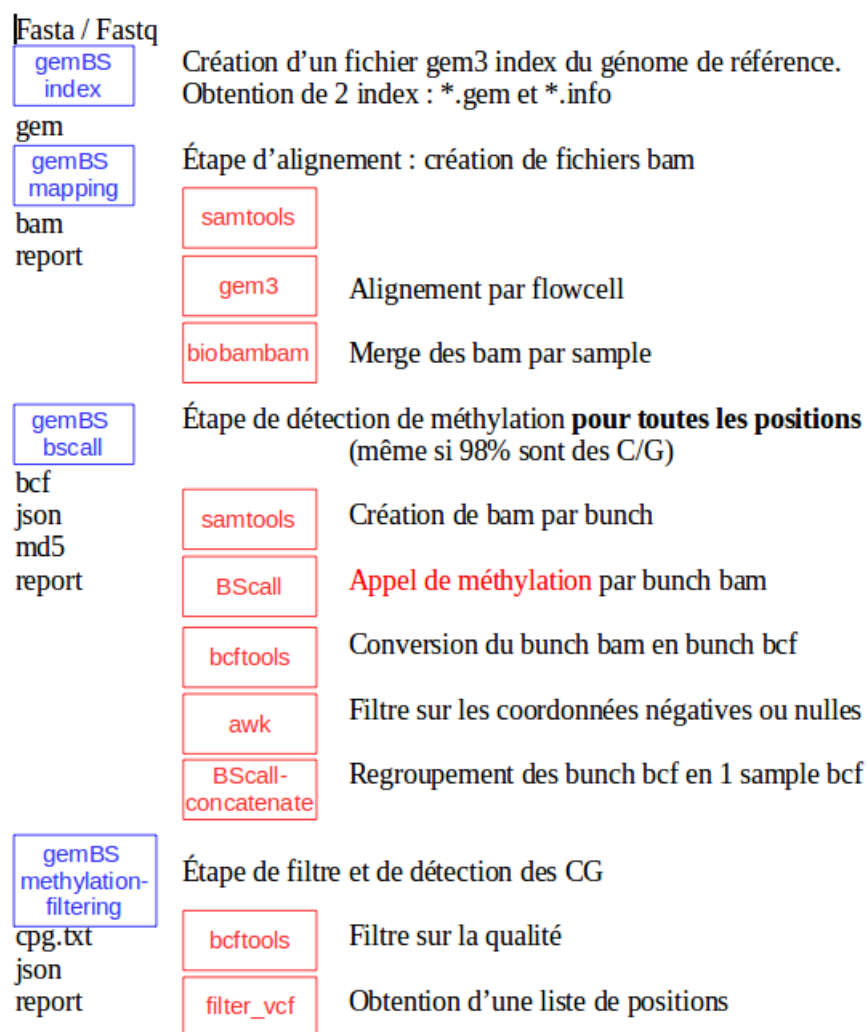


FIGURE 5.8 – Étapes du pipeline gemBS adapté aux données WGBS de poule. Au sein de chaque cadre est notée l'étape du pipeline gemBS concernée ainsi que le logiciel utilisé, le code couleur est le même que dans la figure précédente. Le rôle de chaque étape est décrit à côté des encadrés (en rouge ce qui a été ajouté au pipeline d'origine). Les formats des fichiers de sortie sont notés sous chaque encadré.

Pipelines	Nb lectures initiales	Alignement (%)	Nb lectures alignées	CG détectées	CG filtrées
Toulouse	161 117 566	71,58%	55 986 586	12 467 432	5 733 008
Barcelone	161 117 566	72,84%	58 776 210	14 685 305	5 270 937

Tableau 5.1 – Comparaison des données WGBS de poule. La comparaison des données WGBS poulet se fait entre le pipeline gemBS de Barcelone et le pipeline développé à Toulouse. À partir des mêmes données initiales, le tableau représente le nombre de lectures après alignement et le pourcentage de lecture alignées ainsi que le nombre de CG détectées avant et après les filtres décrits ci-dessus.

GemBS a été construit pour une utilisation sur des données humaines, ce qui nécessite des ajustements pour l'analyse des données de poule. En effet, une des spécificités du génome de la poule est la présence de nombreux *scaffolds* : 874 sur le génome humain (GCA_000001405.27) *versus* 23870 sur le génome de la poule (GCA_000002315.3). La stratégie d'alignement des données avec gemBS se faisant par chromosome, il n'était pas possible de faire de même pour le génome du poulet avec ces trop nombreux *scaffolds*, à cause des ressources informatiques utilisées. Le choix a donc été fait de regrouper ces *scaffolds* par groupe, ou *bunch*, pour permettre l'alignement de nos lectures sur chaque *bunch*. Chaque *bunch* contient une centaine de *scaffolds*. Une fois l'information de méthylation détectée pour chaque fichier, tous les *bunch*, y compris ceux correspondants aux chromosomes, sont fusionnés par individu. La dernière étape est de filtrer les positions pour obtenir un sous-ensemble de CG. La figure 5.8 reprend les étapes du pipeline ainsi que les modifications apportées pour analyser les données. Ce pipeline utilise un cluster de calcul dont le gestionnaire de calcul est SLURM.

Les données WGBS de la poule ont été analysées avec les deux pipelines. Les résultats sont décrits dans le tableau 5.1. À partir du même nombre de lectures pour chaque embryon, le taux d'alignement est proche. L'utilisation de GEM3 permet de détecter un plus grand nombre de lectures que Bismark. Ce dernier est un aligneur de type "3-lettres" qui réalise l'alignement en deux temps : premièrement les C des lectures du brin sens sont complètement convertis en T avant l'alignement, de même que les G sur les lectures du brin anti-sens sont convertis en A. Deuxièmement, en utilisant un aligneur classique, Bowtie qui est intégré à Bismark, l'alignement est alors réalisé sur deux versions modifiées de la séquence du génome de référence (cf partie 1.3.3.1 page 1.3.3.1). La lecture alignée obtenue est alors restaurée pour donner sa séquence d'origine. Dans le cas de GEM3, toutes les étapes de conversion avant et après l'alignement sont réalisées paires de lectures après paires de lectures : un seul alignement est effectué, ce qui évite à la fois la génération de fichiers intermédiaires et diminue le nombre d'étapes du processus. Ensuite, le logiciel BScall détecte les positions de CG et calcule un taux de méthylation. Ce logiciel détecte un nombre plus important de CG que MethylKit. BScall se base sur le modèle de Bis-SNP, logiciel qui permet de détecter les SNP issus de lectures bisulfitées en utilisant une étape de réaligement local [243]. Les outils de détection des variants attribuent un score de qualité de l'alignement généralement faible aux lectures comportant des variants. Réaligner ces lectures permet d'obtenir un score de qualité plus précis, et parfois meilleur [168]. Cependant, BScall ne réalise pas de réaligement local comme Bis-SNP car les variants ne sont pas recherchés. Il peut donc y avoir un bon score de qualité associé à un SNP C/T qui n'est pas une C convertie en T (cf partie 1.3.3.1 page 1.3.3.1). Pour enlever ces positions ambiguës de CG, un second filtre est appliqué sur le score de qualité, ce qui fait chuter le nombre de CG à 10,45 millions en moyenne au lieu des 14,69 millions détectés à partir des fichiers d'alignement. Après la détection des CG, les filtres décrits précédemment (minimum et maximum de couverture, positions communes et suppression de SNP issus du génomique) sont appliqués sur les deux jeux de données. Sensiblement le même nombre de CG sont obtenus : environ 5,7 millions pour le pipeline toulousain et 5,2 millions pour le pipeline barcelonais. Parmi ces positions, 96,3% sont retrouvées dans les deux analyses. Concernant le taux de méthylation, il y a peu de différence dans le profil de méthylation des CG détectés, la majorité des différences étant inférieures à 3% (Figure 5.9). En conclusion, cette comparaison d'analyses de nos données permet de confirmer que les travaux de mise en place d'enchaînements de traitements bioinformatiques pour des analyses de méthylation, dont le pipeline développé pendant la thèse, aboutissent à des résultats comparables, même sans utiliser les mêmes outils.

5.2 Conclusion générale

Le but de cette thèse était de caractériser le méthylome chez la poule par séquençage haut-débit. Deux projets de recherche avec des questions biologiques complémentaires permettaient d'établir les profils de méthylation le long du génome, et le lien avec l'expression, d'abord par un séquençage tout-génome, puis par un séquençage d'une sous-représentation du génome. Dans les deux projets, cette analyse a été réalisée grâce au développement d'un pipeline bioinformatique (Partie 2.3 page 81). Le premier apport de ce travail à la communauté scientifique est donc la disponibilité d'un pipeline d'analyses de données de séquençage haut débit d'ADN après traitement au bisulfite de sodium, optimisé et facile d'utilisation. Ce pipeline est adapté à un cluster de calcul pour pouvoir analyser un grand volume de données avec une gestion optimale des ressources informatiques. Il est optimisé pour faciliter l'usage par tous : une ligne de commande peut exécuter l'ensemble des analyses, mais chaque étape peut s'exécuter indépendamment d'une autre. Plusieurs options définies par des paramètres en entrée permettent d'adapter ce pipeline aux différents types de données de méthylome issues du séquençage à haut-débit.

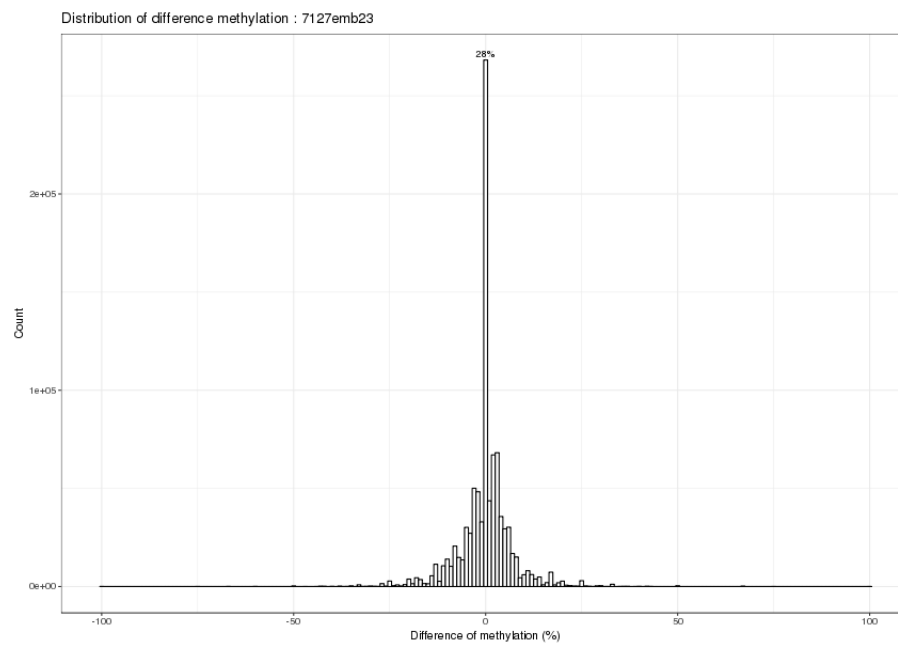


FIGURE 5.9 – Différence du taux de méthylation pour un individu entre les deux pipelines. L'individu est 7127emb23. Les profils de méthylation sont les mêmes pour tous les individus.

La méthylation a été caractérisée dans l'embryon de poule, qui combine plusieurs types de cellules différentes (Partie 3 page 83) et dans l'hypothalamus d'individus adultes (Partie 4 page 133). Seulement trois études WGBS sur le poulet ont été effectuées à ce jour, deux d'entre elles nous ont permis de décrire le paysage de la méthylation de l'ADN chez le poulet dans la revue [158][189][190]. Globalement, le profil de méthylation chez la poule est similaire à ce qui est connu chez les mammifères, avec des niveaux de méthylation plus faibles dans les promoteurs et les îlots CpG. La troisième étude, publiée depuis la publication de notre revue, explore la réponse de la méthylation à l'altitude chez des poulets qui se sont adaptés à cet environnement [191]. À partir d'échantillons sanguins, les auteurs ont montré que la méthylation dans les promoteurs des poulets tibétains est relativement faible, tandis que celle au sein du gène est plus variable, avec deux pics de méthylation observés dans les régions des exons internes et du dernier exon. Le niveau de méthylation dans le promoteur est négativement corrélé à l'expression du gène associé. Cette étude démontre le rôle potentiel de la méthylation de l'ADN, en tant que régulateur épigénétique, en réponse à un changement environnemental. À ce jour, aucune étude d'analyses de méthylome par RRBS chez la poule n'a été publiée.

Les variations de la méthylation montrées dans ce travail de thèse sur le génome entier ou dans une sous-représentation du génome ont révélé que le génome est globalement méthylé, à l'exception de certaines régions qui régulent l'expression des gènes. Leur présence dans le génome peut être caractérisée par un nombre de CG important et un taux de méthylation global faible. Ces îlots CpG sont peu présents dans le génome mais influencent le niveau d'expression des gènes. La majorité des îlots CpG sont présents dans les promoteurs, ce qui fait coïncider leur rôle fonctionnel avec l'initiation de la transcription. Environ 70% des promoteurs annotés sont associés à des CGI chez l'Homme [108]. Dans nos analyses (Partie 3.1 page 83), on observe 57% de promoteurs associés à des CGI détectés à partir de la version 5 du génome de la poule (GCA_000002315.3). Cette plus faible intersection peut provenir de l'annotation fonctionnelle moins complète chez la poule que chez l'Homme. Chez la poule, moins de gènes sont annotés que chez les principales espèces de vertébrés (57373 gènes dans l'annotation 93 de *Ensembl* du génome humain *versus* 24881 dans l'annotation 93 du génome de la poule). Néanmoins, les CGI obtenus dans notre étude sont globalement peu méthylés dans les promoteurs sur les données WGBS (environ 10%, partie 3.1 page 83) et RRBS (environ 4%, partie 4.2 page 137), ce qui confirme ce qui a été observé dans la littérature. Les analyses différentielles de l'effet du sexe sur le méthylome des embryons séquencés en WGBS ont confirmé des caractéristiques déjà établies, comme l'absence d'un phénomène épigénétique permettant la compensation de dose sur les chromosomes sexuels, ou la présence sur le chromosome Z d'une région hyperméthylée, la MHM (Partie 3.1 page 83). De plus, les analyses de la méthylation différentielle entre des animaux contrôles et stressés révèlent une hyperméthylation globale des CG sur le génome, suggérant une réponse de la méthylation en réponse à un stress environnemental (Partie 4.3 page 143).

L'analyse du lien entre la méthylation et la variation du niveau d'expression des gènes a confirmé le caractère complexe de cette relation. Un seul gène différentiellement exprimé et dont le promoteur est différentiellement méthylé entre sexes a été détecté : le gène *H1F0*. Nous avons observé une hyperméthylation de zones introniques associées soit à une régulation positive de l'expression (exemples des gènes *COL6A3* et *KRT14*), soit à une répression (exemple du gène *A1CF*) (Partie 3.1 page 83). Si on sait qu'en général la méthylation du promoteur régule négativement l'expression quand la méthylation du corps du gène la régule positivement, cette régulation peut en effet se révéler plus complexe pour certains gènes ou dans certains tissus [244][245]. Le dispositif EpiBird a également mis en évidence des processus biologiques incluant des gènes différentiellement méthylés entre les croisements réciproques dont sont issus les embryons, comme l'adhésion cellulaire, les processus de développement et la communication cellulaire, qui ont tous un rôle dans le développement embryonnaire. Néanmoins, aucune conclusion biologique concernant le fonctionnement du génome dans le cas de croisements réciproques ne peut être clairement tirée de ces analyses : les niveaux de méthylation des lignées d'origine ne sont pas connus, les analyses concernent des embryons entiers (ne permettant pas de mettre en évidence d'éventuels mécanismes tissu-spécifiques) et nous ne disposons pas de mesures phénotypiques associées, qui, obtenues à partir d'animaux issus des croisements réciproques, pourraient nous permettre de faire le lien entre méthylome embryonnaire et phénotypes adultes. Cependant, le dispositif n'était pas construit pour répondre à ces questions scientifiques : le choix de ces lignées était avant tout de permettre de détecter l'origine parentale des allèles sans confusion. C'est grâce à ce dispositif qu'on a pu mettre en évidence une méthylation allèle spécifique entre les lignées. Grâce à la disponibilité des génomes parentaux, un double génome rassemblant les haplotypes parentaux a permis de détecter l'origine de la méthylation sur nos lectures, et d'éliminer une partie des biais d'alignement [200]. Nous avons ainsi mis en évidence pour la première fois chez la poule le

phénomène de méthylation allèle-spécifique, à une fréquence comparable à ce qui a été observée chez l'Homme (Partie 3.2.4 page 119).

Les premiers résultats des analyses du lien entre la méthylation et la variation de l'expression des gènes soulignent leur association dans les mécanismes biologiques en réponse aux stress, notamment en réponse à l'utilisation d'un aliment pauvre en énergie. Nous avons observé que des gènes différentiellement exprimés entre les groupes contrôles et carencés en aliments énergétiques sont impliqués dans le métabolisme lipidique (DNAJC16, S100A16 et BCAS3) (Partie 4.3 page 4.3) : l'utilisation d'aliments moins énergétiques entraînerait une plus grande mobilisation des réserves lipidiques. Le faible niveau d'expression du gène DNAJC16 en condition FS pourrait entraîner une diminution de l'inflammation dû à la réaction du stress alimentaire, et le fort niveau d'expression de S100A16 en condition FS stimule l'adipogenèse en diminuant l'ostéogenèse. Les données macroscopiques sur les lignées indiquent une diminution du poids corporel en condition FS. Il serait intéressant de regarder les phénotypes, et notamment les quantités de gras et la densité des os entre les deux conditions pour connaître si il existe un lien entre l'expression de ces gènes et la réponse à un régime carencé en énergie. De plus, le gène BCAS3 a un niveau d'expression plus faible en condition FS, et a été détecté comme associé à l'allèle FTO, allèle présent chez les individus avec un risque d'obésité plus important. Ainsi, la mise en évidence de la régulation de l'expression génique positivement corrélée à la méthylation du promoteur pour certains gènes, en cas de stress alimentaire confirme le lien complexe entre ces mécanismes, et une caractérisation de cette variation au sein des caractéristiques géniques fonctionnelles pourrait améliorer la connaissance du lien entre la méthylation et l'expression.

En outre, l'analyse des régions différentiellement méthylées dans le stress à la chaleur a fait ressortir trois gènes (PRRX1, CD93 et SMG6). Seul le gène PRRX1 a déjà été détecté dans la littérature comme impliqué dans la réponse à un stress environnemental. Ce gène intervient dans le mécanisme de l'adipogenèse. Le plus faible niveau d'expression de ce gène en condition HS induit une diminution du taux d'adipocytes. Au sein des lignées, l'exposition à la chaleur entraîne une diminution de poids et du dépôt de tissu adipeux, ce qui est cohérent avec la faible expression du PRRX1 en condition HS. Le faible nombre de gènes différentiels ne permet pas d'explorer de façon plus approfondie le lien avec la méthylation. De plus, les analyses des différentiels de la méthylation entre les 2 lignées sélectionnées sur le critère de consommation résiduelle ont montré que le gène MRC2, impliqué dans la croissance osseuse : la lignée R⁻ a un plus petit tarse que la lignée R⁺ [183]. Ces premiers résultats ont été analysés récemment et sont à approfondir, mais ils laissent à penser que certains gènes sont bien en partie régulés par la méthylation de l'ADN et sont soumis à la variation de l'environnement.

5.3 Perspectives

Au cours du travail de thèse, toutes les données RRBS n'ont pas pu être générées, et leurs analyses sont en lien direct avec ce travail qui caractérise le méthylome par séquençage haut-débit. Des contaminations de dimères d'amorces dans les fragments d'ADN ont nécessité de nouvelles préparations de bibliothèques, et une partie des échantillons ont été re-séquencés. Une perspective majeure sera d'analyser la réponse du méthylome du foie et du sang chez les individus adultes dans la réponse à des stress environnementaux. En plus de l'hypothalamus, le foie maintient l'homéostasie énergétique, il serait donc attendu d'observer, au moins en partie, des changements de méthylation allant dans le même sens que ceux observés dans l'hypothalamus et d'obtenir des gènes impliqués dans les mêmes voies de régulations que celles détectés dans le chapitre 4. La particularité du sang est sa connexion entre tous les tissus de l'organisme. Sa variation de méthylation nous donnerait des éventuelles informations sur les modifications inter-tissulaires et l'implication de gènes associés pourrait révéler leur rôle dans la tolérance aux stress environnementaux [246]. Dans un second temps, réaliser une méta-analyse de ces trois prélèvements, sur la variabilité de leur méthylation et le lien avec les variations d'expression des gènes permettrait d'avoir une vision globale de la réponse du méthylome à différents stress environnementaux chez la poule. À ces données de méthylation et d'expression peuvent s'ajouter les informations phénotypiques, générées par ailleurs dans le programme ChickStress, comme par exemple des mesures métaboliques, ce qui permettrait de mieux décrypter la complexité des réponses physiologiques aux variations de l'environnement.

Cependant, la construction de ce dispositif, très complexe (plusieurs types de données pour plusieurs types d'analyses dans le but de mettre en évidence un phénomène biologique), doit notamment prendre en compte la

variabilité entre les individus. Dans un même groupe étudié, avec les mêmes conditions (même tissu, âge, lignée, ...), la variabilité propre à chaque individu peut révéler un méthylome variable entre individus. Une analyse approfondie des résultats obtenus en transcriptome et en méthylome en fonction du fond génétique apporterait probablement des éléments intéressants sur le contrôle génétique de la variabilité épigénétique. Cette analyse pourrait même gagner en précision en analysant, comme nous l'avons effectué dans le chapitre 3 page 83, les phénomènes allèles-spécifiques potentiellement impliqués. Une analyse intéressante et originale de ces données consisterait à observer la variabilité inter-individuelle des niveaux de méthylation des dinucléotides CG : des régions à faible variabilité peuvent être le reflet de fonctions importantes dans le fonctionnement normal de l'individu ; des régions fortement variables seraient potentiellement soumises à des régulations épigénétiques plus fines, en réponse aux variations de l'environnement.

Concernant le premier programme, la construction du dispositif EpiBird a été réfléchi pour diminuer au maximum la variabilité entre les individus d'une même lignée et maximiser la variabilité entre les deux lignées dans le but de détecter l'empreinte génomique parentale. Ce dispositif a également été utilisé pour mettre en évidence l'édition de l'ARN [179]. C'est un phénomène qui modifie l'ARN, et potentiellement la séquence en acides aminés, entraînant des modifications similaires à des situations correspondant à des mutations génomiques. Une étude récente a montré l'influence des allèles dans l'édition de l'ARN entre les tissus et les espèces (Homme et souris) [247]. Grâce aux données du projet EpiBird, il serait intéressant d'analyser ce phénomène allèle-spécifique chez le poulet.

Les données générées dans les deux projets scientifiques abordés au cours de ce travail avaient pour but de caractériser le méthylome de la poule, à l'échelle du génome entier ou à travers une sous-représentation du génome. L'utilisation de 2 types de données ainsi que les différentes méthodes utilisées ont permis de croiser les données pour mettre en évidence les caractéristiques du méthylome chez la poule, qui sont retrouvées chez les vertébrés. L'amélioration de l'annotation grâce aux apports technologiques permet de caractériser toujours plus finement les régions fonctionnelles du génome. Un exemple est la corrélation négative de la méthylation dans l'intron 1 des gènes avec l'expression, association qui n'est pas tissu-spécifique et est présente chez les vertébrés [245]. La très prochaine version du nouveau génome de la poule (GCA_000002315.5) devrait permettre d'obtenir une annotation du génome plus complète, et ainsi détecter plus de régions du génome annotées pour les analyses décrites dans cette thèse.

L'utilisation de deux techniques haut-débit a généré de gros volumes de données pour répondre aux questions biologiques. Le tri de cette masse de données passe par un traitement bioinformatique, qui détecte le phénomène étudié tout en évitant au maximum les biais inhérents aux données. Bien que nos choix aient été validés quant aux méthodes utilisées, de nouvelles méthodes sont aujourd'hui en développement et impliquent une mesure simultanée de multiples caractéristiques génomiques dans les cellules individuelles. Par exemple, le NOME-seq (*Nucleosome Occupancy and Methylome sequencing*) intègre les informations de méthylation et de positionnement du nucléosome pour permettre de connaître les structures de la chromatine qui régulent l'expression des gènes [248]. Une autre technologie, dite de troisième génération, le Nanopore (nanoporetech.com), permet de détecter directement si le C est méthylé ou non [249]. En effet, chaque molécule d'ADN est immergée dans une solution conductrice et passe dans un pore à l'échelle nanométrique. Le changement du courant électrique spécifique à une base de l'ADN permet de lire directement la séquence d'ADN (séquençage classique). L'effet de la méthylation sur le courant ionique peut être observé en comparant, pour la même séquence, les niveaux de courant observés à partir de lectures d'ADN méthylé avec ceux issus de lectures non méthylées. Le groupement méthyle du C entraîne une augmentation de courant par rapport à un C non méthylé. Cette technologie permet de séquencer de grands fragments et produit à la fois la séquence du génome et du méthylome, sans traitement au bisulfite, ce qui élimine les biais connus du *bisulfite-sequencing*. Ces nouvelles méthodes et ces progrès récents ont été discutés à l'échelle européenne dans le cadre du congrès ECCB, *European Conference on Computational Biology*, du 9 au 12 septembre 2018.

Néanmoins, la méthylation de l'ADN n'est pas le seul facteur qui peut affecter la transcription des gènes. Par conséquent, étudier d'autres mécanismes, et notamment d'autres marques épigénétiques, peut être intéressant pour éclairer ces mécanismes sous-jacents impliqués dans l'adaptation à différents environnements. Les collaborateurs de la revue ont caractérisé les modifications d'histones dans l'acclimatation embryonnaire à la chaleur chez le poulet de chair [250]. Comme pour la méthylation de l'ADN, les autres marques épigénétiques bénéficient de techniques de séquençage haut-débit et des méta-analyses apparaissent pour explorer la diversité

des mécanismes moléculaires dans de nombreuses espèces. En plus de l'aspect fondamental de cette recherche, l'application plus concrète de ces connaissances peut s'appliquer aux systèmes d'élevage dans le but d'améliorer les caractères de production. Avec l'évolution du climat mondial, l'obtention d'animaux plus robustes, plus performants et surtout plus adaptés à leur environnement est un challenge pour les années à venir.

Annexe A

Revue dans *Productions Animales*

Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire

Marjorie Mersch^{1†}, Sarah-Anne David^{2†}, Anaïs Vitorino Carvalho^{2†}, Sylvain Foissac¹, Anne Collin², Frédérique Pitel^{1*}, Vincent Coustham^{2*}

¹ GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet Tolosan, France

² BOA, INRA, 37380 Nouzilly, France

*Courriel : frederique.pitel@inra.fr; vincent.coustham@inra.fr

Résumé

Comme toutes les espèces de rente, les oiseaux d'élevage ont été principalement sélectionnés génétiquement afin d'augmenter leurs performances. Cependant, de nombreuses études ont montré que le phénotype d'un individu n'était pas le simple résultat de son patrimoine génétique mais résultait également de son environnement. En effet, l'expression des gènes est influencée par l'environnement (température, nutrition, etc.) notamment via des changements de marques épigénétiques qui jouent ainsi un rôle dans la plasticité des phénotypes des animaux. Aujourd'hui, grâce aux techniques de séquençage à haut-débit devenues suffisamment abordables, il est possible d'étudier des marques épigénétiques à l'échelle du génome. Chez les oiseaux, un nombre croissant d'études utilisent ces techniques pour mieux comprendre les mécanismes épigénétiques impliqués dans la régulation de l'expression des gènes, influençant des processus tels que l'immunité ou l'adaptation à l'environnement. Dans cette synthèse, nous décrivons en quoi ces technologies haut-débit ont contribué à enrichir les connaissances sur l'épigénome aviaire et plus particulièrement celui du poulet, en nous concentrant sur les deux types de modifications épigénétiques les plus étudiées, la méthylation de l'ADN et les modifications post-traductionnelles des histones. Nous présentons également les concepts nécessaires à la bonne conception et réalisation des analyses haut-débit des épigénomes aviaires. En plus des connaissances fondamentales fortement attendues par la communauté scientifique, une meilleure compréhension des mécanismes épigénétiques à l'origine de la régulation de l'expression génique en réponse aux modifications environnementales chez l'oiseau favoriserait le développement d'une production avicole durable.

Abstract

New insights on the avian epigenome from the high-throughput sequencing experiments

Over the years, farmed birds have been genetically selected to improve some performance traits. However, many studies have shown that phenotype not only depends on the genetic but also of the environment. Gene expression is also influenced by the environment, through epigenetic changes that may alter the phenotypes of the animals. Recently, since high-throughput sequencing techniques are extremely developing and become less expensive, a growing number of studies took advantage of them to gain insights into the epigenetic mechanisms of gene regulation in birds. Here we review the current gain of knowledge on the chicken epigenome – focusing on the two most studied epigenetic modifications: DNA methylation and histones post-translational modifications - based on recent data obtained with high-throughput sequencing techniques. We also provide advices

about designing and performing analyses to further explore avian epigenomes. A better understanding of the molecular mechanisms underlying the epigenetic involvement in the bird phenotypes may provide new markers that should undoubtedly contribute to improve the poultry production.

Chapeau

Le phénotype résulte de l'expression du génome, de l'environnement et de leur interaction, en particulier via l'épigénome. Alors que l'oiseau est un modèle largement étudié en biologie du développement, en immunologie et en microbiologie, peu d'informations sont disponibles sur son épigénome. Dans cette revue, nous synthétisons les connaissances concernant l'épigénome aviaire, principalement celui du poulet, et présentons des éléments pour mener à bien son analyse.

Introduction

L'oiseau est un modèle de choix pour l'étude de la biologie du développement (en particulier l'embryogenèse), l'immunologie et la microbiologie. C'est en particulier le cas de la poule. Il s'agit également d'une source majeure de protéines par sa production de viande et d'œufs. Au cours des cinquante dernières années, la sélection génétique a permis un important accroissement de la production avicole grâce notamment à l'obtention d'animaux présentant une croissance rapide. Cependant, des analyses mettant en regard des caractères phénotypiques et des études génomiques ont montré que la variation génétique n'est pas le seul facteur contribuant à la variabilité phénotypique observée (Manolio et al. 2009). Un nombre croissant d'études montrent que l'épigénétique contribue également à la diversité phénotypique des animaux, en interaction avec leur environnement (Triantaphyllopoulos, Ikononopoulos, and Bannister 2016). Ainsi, il a été suggéré que la modulation épigénétique de l'expression des gènes pouvait avoir des impacts à long-terme sur le phénotype des poulets, notamment sur leurs comportements (Verhulst et al. 2016), leur capacité d'adaptation (Kisliouk and Meiri 2009) et sur leurs performances (J. Li et al. 2015). De plus, les techniques de séquençage haut-débit (SHD) ont permis de générer de nouvelles connaissances concernant l'épigénome du poulet.

Dans cette revue adaptée de David et al. 2017 (David, Mersch, et al. 2017), nous présenterons le récent gain de connaissances sur l'épigénome du poulet en nous concentrant sur deux types de marques épigénétiques : la méthylation de l'ADN et les modifications post-traductionnelles des histones (MPTH). Nous nous intéresserons également aux similitudes des profils épigénétiques du poulet avec ceux observés dans d'autres espèces, notamment chez les mammifères. Enfin, nous évoquerons les recommandations à suivre et les pièges à éviter pour la bonne conduite d'études génomiques sur les espèces aviaires.

1. Les marques épigénétiques

Dans les années 1950, Waddington a défini un caractère épigénétique comme « un phénotype stable héréditaire résultant de changements dans un chromosome sans altération dans la séquence d'ADN » (Berger et al. 2009). De nos jours, l'épigénétique est communément définie comme l'ensemble des mécanismes moléculaires impliqués dans la régulation de l'expression des gènes qui sont réversibles et héréditaires (par mitose et potentiellement par méiose) sans altération de la séquence d'ADN. Il a été montré que les modifications épigénétiques chez les eucaryotes peuvent

affecter l'expression des gènes sous l'influence de l'environnement (température, polluants, etc.) (Feil and Fraga 2012).

Il existe de nombreux types de facteurs épigénétiques agissant à différents niveaux de l'expression génique : la méthylation de l'ADN, les MPTH, les variants d'histones ou encore les ARN non-codants (Figure 1). Bien que ces deux dernières modifications peuvent également jouer un rôle majeur dans la variabilité phénotypique (Roundtree et al. 2017), peu de connaissances existent à l'heure actuelle chez le poulet et elles ne seront, par conséquent, pas traitées dans cette revue.

1.1. La méthylation de l'ADN

La méthylation de l'ADN reste actuellement la marque épigénétique la plus étudiée (Bock 2012). Chez les vertébrés, cette marque cible principalement les résidus cytosine à la position 5 du carbone (5mC, (Gardiner-Garden and Frommer 1987)) dans un contexte de dinucléotides CpG. La distribution de la 5mC à travers le génome n'est pas aléatoire (Gardiner-Garden and Frommer 1987). En effet, les dinucléotides CpG ont tendance à être regroupés dans des régions nommées des îlots CpG qui sont souvent hypométhylés (Bock 2012; Gardiner-Garden and Frommer 1987). La présence de 5mC dans les promoteurs des gènes est généralement associée à la répression de la transcription (Bird 1992), alors que sa présence dans le corps des gènes est plutôt favorable à l'expression génique (Jones 2012). Chez les jumeaux monozygotes humains, la 5mC diffère dans plusieurs tissus, suggérant que cette marque contribue à l'expression différentielle de gènes liés à l'histoire de vie de l'individu (Busche et al. 2015). Il existe une corrélation entre la méthylation de l'ADN et la variation des phénotypes chez plusieurs espèces végétales et animales (Feil and Fraga 2012). Par exemple, la couleur du pelage des souris agoutis est régulée par l'expression du gène *ASIP*. L'ingestion d'aliments présentant une forte teneur en donneurs de groupes méthyles (tels que l'acide folique, la vitamine B12, la bêtaïne et la choline) par ces souris agoutis gestantes induit un assombrissement de la couleur du pelage des petits expliqué par la méthylation différentielle d'un rétro-transposon en amont du gène *ASIP* (Wolff et al. 1998). La 5mC n'est pas la seule modification du cinquième carbone des résidus cytosine potentiellement impliquée dans la régulation de la transcription. En effet, de récentes études ont montré que la 5-hydroxyméthylcytosine (5hmC), proposée comme un intermédiaire de déméthylation de l'ADN, influencerait également l'expression de gènes (Ponnaluri et al. 2017).

1.2. Les modifications post-traductionnelles des histones

L'autre catégorie principale de modifications épigénétiques est constituée des MPTH. Chez les eucaryotes, les histones sont des protéines qui conditionnent l'ADN en unités structurales appelées nucléosomes. Les histones H2A, H2B, H3 et H4 forment un noyau enveloppé par 147 paires de bases d'ADN, l'ensemble formant une unité structurale appelé nucléosome (Kornberg and Lorch 1999). Les histones possèdent deux régions : le cœur et la queue. Ce sont les cœurs des histones qui constituent les nucléosomes. Les queues, elles, restent accessibles en dehors de la structure nucléosomale (Figure 1). Les acides aminés principalement situés au niveau de la queue, mais aussi parfois du noyau, peuvent être la cible de modifications covalentes : méthylation, acétylation, ubiquitination, phosphorylation, sumoylation, etc. (Bannister and Kouzarides 2011). L'une des fonctions des MPTH est la régulation structurale des nucléosomes (dont l'interaction entre les histones et la double hélice de l'ADN) et par conséquent de la conformation de la chromatine, qui va être plus ou moins condensée et donc plus ou moins accessible à la machinerie transcriptionnelle (Tessarz and Kouzarides 2014). Ces MPTH jouent également sur le recrutement de protéines sur l'ADN (notamment des facteurs de transcription) et modulent donc directement la transcription des gènes, que ce soit pour la favoriser ou la réprimer (Kouzarides 2007). Par exemple, l'acétylation de la lysine 27 sur l'histone H3 (H3K27ac) et la tri-méthylation de la lysine 4 sur l'histone H3

(H3K4me3) sont communément dites « activatrices » car induisent l'ouverture de la chromatine au niveau des promoteurs et donc favorisent l'expression des gènes (T. Zhang, Cooper, and Brockdorff 2015). En revanche, la triméthylation des lysines 9, 36 et 27 de l'histone H3 (H3K9me3, H3K36me3, H3K27me3) sont dites « répressives » car généralement associées à l'absence d'expression du gène. La présence d'un type de MPTH ne prédit pas nécessairement l'état d'expression d'un gène. En effet, des modèles combinatoires de différents profils de MPTH ont été proposés pour définir des états chromatinien spécifiques (Baker 2011).

2. L'étude de l'épigénome

Il existe de nombreuses techniques pour analyser la présence de modifications épigénétiques (Kurdyukov and Bullock 2016), avec des échelles très variables. Dans cette revue, nous nous focaliserons sur les analyses tout-génome nécessitant le recours au SHD.

2.1. L'étude de la méthylation de l'ADN

A ce jour, la méthode de référence pour l'étude de la 5mC est la conversion de l'ADN génomique par le bisulfite de sodium (Figure 2 et Tableau 1), qui désamine les cytosines en uraciles. Cette désamination est bloquée par la présence d'un groupement méthyle, permettant ainsi de localiser des cytosines méthylées par comparaison au génome de référence (Frommer et al. 1992; Mardis 2008). En effet, une cytosine non convertie après traitement au bisulfite suggère qu'elle était méthylée dans l'échantillon de départ, alors que sa conversion en thymine suggère qu'elle n'était pas méthylée. Une analyse à l'échelle du génome entier peut ainsi être réalisée quand la conversion au bisulfite est combinée au SHD (*Whole Genome Bisulfite Sequencing* - WGBS ou *MethylC-seq* (Urich et al. 2015), Figure 2 et Tableau 1) et que le génome de référence est disponible. Plus récemment, le séquençage des fractions riches en CpG, donc enrichies en îlots CpG, a été développé en introduisant une étape de digestion enzymatique grâce à l'enzyme *MspI*, dont le site de restriction est CCGG (*Reduced Representation Bisulfite Sequencing* - RRBS) (Harris et al. 2010). Cette technique est considérée comme plus économique car elle se focalise principalement sur l'étude des séquences promotrices qui sont généralement riches en CpG, aux dépens des CpG plus isolées ainsi souvent exclues de l'analyse (Gu et al. 2011). L'utilisation de différentes combinaisons d'enzymes peut augmenter la portion du génome étudiée, bien que celle-ci reste une sous-représentation du génome. Ainsi, en fonction de la question posée et des moyens financiers disponibles, il peut être conseillé d'utiliser l'une ou l'autre de ces méthodes. Ces deux techniques nécessitent un ADN de très haute qualité pour obtenir une conversion efficace (Ziller et al. 2015). Une autre limitation de ces méthodes est que la conversion au bisulfite ne permet pas de distinguer la 5mC des 5hmC (Globisch et al. 2010). Pour évaluer la part de la 5hmC dans le signal total, il faut effectuer une étape d'oxydation de l'ADN et comparer les résultats à une expérience sans l'étape d'oxydation (correspondant donc aux fractions 5mC et 5hmC) (Hahn et al. 2015). Cette étape d'oxydation de l'ADN n'est pas systématiquement effectuée dans les protocoles classiques d'étude de la 5mC, car l'oxydation a tendance à endommager l'ADN et la proportion de 5hmC est souvent résiduelle dans la plupart des tissus (Globisch et al. 2010). D'autres protocoles d'étude de la 5mC n'utilisant pas la conversion au bisulfite et se basant par exemple sur l'utilisation d'enzymes thermosensibles sont également utilisés (Tableau 1). Chaque technique présente ses propres biais qu'il est important de connaître afin de choisir le bon outil et d'interpréter correctement les résultats (Bock 2012).

Il est probable qu'à l'avenir les analyses de méthylation haut-débit à l'échelle du génome puissent se passer de la conversion au bisulfite de sodium et de l'étape d'amplification de l'ADN grâce par exemple à la technologie Nanopore, en cours de développement, qui pourrait permettre de lire

directement, en une seule fois, le génome et le méthylome d'un même échantillon (Simpson et al. 2017).

2.2. L'étude des modifications post-traductionnelles des histones

Pour étudier les marques d'histones, la méthode de référence est actuellement l'immunoprécipitation de la chromatine suivie de SHD (ChIP-seq, Figure 2 et Tableau 1). Cette technique vise à identifier les lieux d'association de l'ADN avec des protéines (facteurs de transcription, histones, etc.), par séquençage de fragments d'ADN immunoprécipités en utilisant des anticorps spécifiques aux épitopes d'intérêt (Barski et al. 2007; Johnson et al. 2007; Robertson et al. 2007). Ainsi, par l'utilisation d'anticorps spécifiques des MPTH, il est possible d'identifier des séquences d'ADN qui leur sont associées. Pour ces études, la chromatine extraite peut être soit fixée (protéines liées de façon covalente à l'ADN) et fragmentée par sonication (X-ChIP-seq), soit native (ou non-fixée) et digérée par une enzyme (N-ChIP-seq) (David, Piégu, et al. 2017). Le choix de la méthode dépend de la nature du facteur étudié et du tissu (David, Piégu, et al. 2017). Une étape de digestion enzymatique par exonucléase peut être ajoutée au protocole de X-ChIP-seq pour réduire la taille des fragments séquencés (ChIP-exo, (Rhee and Pugh 2011)) et ainsi améliorer sa résolution. Ainsi, l'état de la chromatine (fixée ou non), la fragmentation et l'affinité de l'anticorps sont des points qui peuvent influencer les résultats obtenus. Il est donc crucial de mettre au point les conditions expérimentales avec soin et de prendre en compte le choix de la méthode lors des analyses de données (David, Piégu, et al. 2017).

3. La conception d'études épigénomiques

Une particularité essentielle des études haut-débit est qu'en raison du volume de données générées elles nécessitent une forte capacité d'analyse bioinformatique. Ainsi, pour faciliter ces analyses et obtenir des résultats les plus fiables possibles, il est important d'intégrer des bioinformaticiens et des biostatisticiens dès la conception du projet (Rigail et al. 2016). Pour guider les analyses, de nombreux consortiums ont défini des recommandations.

3.1. Les recommandations pour l'analyse du séquençage haut-débit

Au cours des deux dernières décennies, les techniques de SHD se sont démocratisées grâce notamment à la baisse de leur coût, entraînant ainsi une meilleure compréhension des transcriptomes et des épigénomes. Afin d'homogénéiser les résultats, plusieurs consortiums sont apparus pour définir des recommandations sur les protocoles basés sur le SHD. Ces recommandations concernent l'élaboration, l'application et l'analyse de résultats de ces protocoles et ont été émises principalement à partir des données obtenues chez les mammifères, afin d'encadrer toute procédure touchant au SHD et ainsi faciliter la comparaison des résultats entre les études. Ces différentes directives sont issues de projets variés comme le programme ENCODE (ENCyclopedia Of DNA Elements) (ENCODE Project Consortium 2012), le consortium NIH (National Institute of Health) Roadmap Epigenomics Project Mapping Consortium (Bernstein et al. 2010), le consortium IHEC (International Human Epigenome Consortium) (Bae 2013) et l'application d'ENCODE aux espèces modèles (modENCODE) (Celniker et al. 2009). Depuis 2015, la volonté de mieux comprendre la diversité génétique, épigénétique et phénotypique des espèces d'élevage comme le poulet a conduit à la création d'un consortium appelé FAANG (Functional Annotation of ANimal Genomes) (Andersson et al. 2015). Au-delà de son objectif de doter les génomes animaux d'annotations fonctionnelles de haute qualité – épigénomes compris –, FAANG contribue aussi à la recherche médicale et agronomique en normalisant les tests de base et les protocoles expérimentaux et en établissant une infrastructure pour l'analyse des données haut-débit.

Les recommandations émises par ces consortiums concernent notamment le nombre de réplicats biologiques, mais également la profondeur de séquençage nécessaire à l'obtention de résultats de SHD vérifiables. La profondeur de séquençage est définie comme « le nombre moyen de fois qu'un nucléotide particulier (une position définie sur le génome) est représenté dans une collection de séquences brutes aléatoires » (Stein 2001) et s'exprime en « fois » (X). La première étude tout-génome de la méthylation de l'ADN a été publiée en 2008 sur *Arabidopsis thaliana* avec une profondeur de séquençage d'environ 20 X (Cokus et al. 2008). Rapidement, d'autres études ont adapté cette méthode chez différentes espèces, en particulier chez l'homme (Lister et al. 2009) et la souris (Smith et al. 2009). Pour le WGBS, il est ainsi recommandé par le Epigenomics Roadmap de séquencer à un minimum de 5 X de profondeur alors que pour le RRBS, une profondeur de 10 X est préférée (Doherty and Couldrey 2014). La quantité de séquences nécessaires à une analyse RRBS dépend de la proportion d'ADN sélectionnée et donc du profil de l'enzyme de restriction utilisée : pour une même quantité totale de lectures, la profondeur en chaque site observé sera d'autant plus élevée que la représentation du génome est réduite. À notre connaissance, aucune étude de RRBS chez le poulet n'a été publiée ; par conséquent, le profil de l'enzyme de restriction *MspI* doit être soigneusement examiné en amont pour estimer la quantité de séquences nécessaires chez cette espèce.

La première étude tout-génome de MPTH sur des cellules humaines date de 2007 (Barski et al. 2007). Depuis, le consortium ENCODE a publié les lignes directrices pour obtenir des résultats correctement interprétables d'expériences de ChIP-seq : la qualité de l'anticorps, les contrôles utilisés, le nombre de réplicats biologiques et le nombre de lectures uniques obtenues après le séquençage (Landt et al. 2012). Le consortium ENCODE recommande 20 à 40 millions de lectures uniques pour l'analyse ChIP-seq en fonction du type de marque étudiée (Landt et al. 2012). Pour la marque H3K27me3, présentant des pics larges, une couverture élevée de 40-50 millions de lectures alignées uniques est recommandée pour identifier des pics chez l'humain (Jung et al. 2014). (David, Piégu, et al. 2017) Bien que ces recommandations aient été initialement définies pour les espèces de mammifères, elles devraient être suivies pour les études de ChIP-seq chez les oiseaux. En effet, pour H3K27me3, 40 millions de lectures alignées de façon unique semblent être suffisantes chez les oiseaux (David, Piégu, et al. 2017). Il est important de noter que plus l'immunoprécipitation est efficace, moins il sera nécessaire d'augmenter la profondeur de séquençage (David, Piégu, et al. 2017).

3.2. Le plan expérimental

Le SHD est maintenant bien documenté chez de nombreuses espèces (dont l'humain, la souris, la drosophile et les plantes), nous permettant ainsi de définir des considérations communes à l'étude d'un épigénome quelle que soit l'espèce, et donc également valables pour les oiseaux. Ainsi, de nombreux facteurs peuvent influencer l'épigénome comme l'âge des individus ou encore leur histoire de vie (conditions d'élevage, nutrition, etc.). Il est également important de connaître l'histoire parentale dont l'âge des reproducteurs et les conditions de transport, de conservation des œufs et d'éclosion. Enfin, idéalement, il est intéressant de contrôler la variabilité génétique des individus étudiés, pour maximiser les chances d'observer des différences épigénétiques significatives. L'utilisation de lignées fortement consanguines peut permettre d'étudier les modifications épigénétiques dans une situation où la variation génétique est nulle ou presque, et pourrait donc faciliter l'interprétation des résultats. Ainsi, biologiquement on élimine la variabilité épigénétique sous le contrôle de la génétique, et techniquement, pour la méthylation de l'ADN où le traitement au bisulfite entraîne un polymorphisme apparent, on élimine les taux de méthylation artéfactuels dus à un polymorphisme ADN de type C/T (Daxinger and Whitelaw 2010). En plus des biais techniques liés aux protocoles de biologie moléculaire précédemment évoqués, tous ces

aspects doivent donc être pris en compte dans les analyses statistiques à défaut d'être strictement contrôlés.

4. L'analyse bioinformatique

Souvent sous-estimée, l'analyse des données issues de SHD est un autre élément critique dans les études épigénomiques. Elle nécessite des compétences analytiques particulières ainsi qu'une forte capacité de calcul et de stockage des données.

4.1. Les compétences analytiques

Actuellement, les plateformes de SHD proposent généralement l'analyse des données générées. Cependant, la réalisation d'une analyse complète de données épigénétiques reste difficile car elle nécessite une version correctement annotée du génome ainsi que des outils informatiques dont les mises à jour sont fréquentes (Figure 3). Ainsi, la première étape est de vérifier la disponibilité du génome ainsi que son annotation. Ensuite, il faut prendre en compte l'évolution rapide des outils bioinformatiques qui peut entraîner l'utilisation de nouveaux formats de stockage et des algorithmes de traitement au cours du projet. Par exemple, les approches de référence jusque-là utilisées pour l'alignement (identification de la position d'une séquence sur le génome de référence) comme les approches BLAST/BLAT se sont révélés inefficaces pour traiter les positions de milliards de courtes lectures sur le génome entier. Ainsi, des méthodes spécifiques ont dû être développées pour répondre à ce type d'alignement (Langmead et al. 2009; H. Li and Durbin 2009). Des outils comme Bowtie (Langmead et al. 2009), BWA (H. Li and Durbin 2009) et SOAP2 (R. Li et al. 2009) présentent chacun des paramètres spécifiques (par exemple le nombre de bases non identiques à la référence autorisées pour valider l'alignement) qui impactent significativement l'alignement des lectures et donc l'identification des zones d'intérêt (Han and He 2016). Ces spécificités ont été intégrées dans plusieurs pipelines mis au point pour étudier les données issues de SHD comme le pipeline d'analyse de WGBS du consortium ENCODE (The ENCODE Project Consortium 2017) ou le pipeline d'analyse de ChIP-Seq du BLUEPRINT (Epigenome Blueprint 2017). Des normes similaires sont attendues pour les espèces aviaires dans le cadre du consortium FAANG (Andersson et al. 2015).

4.2. Les capacités de calcul et de mémoire

Deux autres facteurs importants concernent la capacité de calcul et de mémoire pour exécuter les pipelines d'analyses bioinformatiques, et l'espace du disque dur pour stocker les données et les résultats. Les installations de bioinformatique fournissent généralement des supports informatiques pour mettre à disposition des capacités de calcul suffisantes (par exemple, Genotoul-Bioinfo, plateforme bioinformatique du génopole de Toulouse - <http://bioinfo.genotoul.fr/>). Tout comme le protocole expérimental et le budget d'un projet donné, le volume de séquences brutes obtenues doit être défini en amont. Alors qu'on observe une réduction drastique de la taille des données tout au long de l'analyse bioinformatique, les opérations initiales comme le nettoyage des données et l'alignement des lectures issues de SHD produisent des fichiers volumineux. A titre d'exemple, l'ensemble des fichiers générés par une analyse WGBS d'environ 10 X de profondeur chez le poulet (données brutes, intermédiaires et finales) représente près de 35 Go par échantillon. Chez le même modèle, les analyses de ChIP-seq, elles, nécessitent environ 10 Go de stockage. Elles comprennent aussi bien les séquences des fractions contrôles et immunoprécipitées (fichiers fastq.gz compressée initiaux - environ 1 à 2 Go) que les fichiers de coordonnées des pics (500 Ko). Ainsi il est nécessaire de prévoir des capacités de stockage adéquates. Les fichiers intermédiaires, souvent volumineux, peuvent être éliminés à la fin des analyses mais sont souvent gardés car un pipeline est rarement utilisé une seule fois. Par exemple, chaque mise à jour d'un logiciel est

susceptible d'affecter les résultats et donc de nécessiter une nouvelle exécution de toute l'analyse. De plus, la comparaison des résultats générés par les différents pipelines fait partie du contrôle de la qualité des données. Il est donc fortement conseillé d'anticiper un volume de stockage important pour toute analyse impliquant du SHD.

En outre, les besoins de stockage peuvent s'avérer supérieurs dans les cas des espèces de rente car les analyses impliquent souvent de nombreuses optimisations (test de différents paramètres et de logiciels qui ne sont pas développés pour l'étude de ces génomes). En effet, certains outils d'analyse sont optimisés pour l'étude d'un génome constitué de 22 + XY chromosomes humains et non pas des 32 + ZW chromosomes et 524 « scaffolds » que contient l'assemblage actuel du génome du poulet (GRCg6a ; GCA_000002315.4). L'adaptation de ces protocoles aux organismes non-modèles requiert donc souvent quelques ajustements.

Enfin, l'étude des marques épigénétique sera d'autant plus pertinente et informative que le génome de référence de l'espèce étudiée disposera d'un assemblage de son génome de bonne qualité, et bien annoté.

5. L'épigénome du poulet

5.1. Le génome du poulet

Le génome du poulet est environ 2,5 fois plus petit que le génome humain (1200 Mb contre 3200 Mb). Il est composé de 32 paires d'autosomes et de deux chromosomes sexuels (Z et W). La plupart des autosomes (29 paires au total) sont relativement petits et de taille uniforme, et sont appelés « microchromosomes ».

Le génome du poulet fut le premier génome aviaire, et le premier génome d'une espèce de rente, à avoir été séquencé (International Chicken Genome Sequencing Consortium. 2004). L'assemblage du génome de *Gallus gallus* a été obtenu à partir d'une seule poule de race "Red Jungle Fowl". La dernière version de l'assemblage du génome du poulet (GRCg6a ; GCA_000002315.4), d'une taille de 1,1 Gb, a été publiée en 2018 et a été obtenue grâce à la technologie de lecture longue (PacBio). Cette sixième version du génome a été améliorée de façon notable, avec une réduction du nombre de « scaffolds » (lectures non identifiées dans le génome de référence) qui est passé de 23870 (version 5) à 525. [63] : env

Environ 38 % des îlots CpG sont conservés entre l'homme et le poulet (International Chicken Genome Sequencing Consortium. 2004). Chez cette espèce, les îlots CpG sont principalement localisés sur des microchromosomes présentant une forte densité de gènes (McQueen et al. 1996).

5.2. La méthylation de l'ADN chez le poulet

La première analyse de méthylation de l'ADN dans le foie et les tissus musculaires du poulet a été publiée en 2011 (Q. Li et al. 2011). La technique du MeDIP-seq était utilisée. Cette étude a permis de mettre en évidence une diminution du niveau de méthylation de l'ADN dans les régions promotrices, en particulier au niveau du site d'initiation de la transcription (TSS), et une augmentation dans des régions du corps du gène (Q. Li et al. 2011). Ces observations sont cohérentes avec la localisation des îlots CpG qui sont principalement dans les promoteurs, et confirmées dans une étude ultérieure basée sur la technique de MeDIP-seq (Hu et al. 2013). De légères variations de méthylation ont été observées le long des gènes (Q. Li et al. 2011). La densité de méthylation est plus élevée dans les introns que dans les régions transcrites non traduites (UTRs) et les exons (Hu et al. 2013). Les niveaux de méthylation des régions promotrices sont négativement corrélés avec l'expression des gènes associés. On retrouve donc le même type de corrélation entre niveaux d'expression et méthylation que celle observée chez l'humain (Suzuki and Bird 2008). Le niveau de méthylation varie entre les tissus, l'ADN hépatique étant moins méthylé

que celui du muscle (Q. Li et al. 2011). Malgré une densité d'éléments répétés inférieure chez le poulet (11 %) que chez les mammifères (40 – 50 %) (International Chicken Genome Sequencing Consortium. 2004), les profils de méthylation sont similaires avec environ 60 % des séquences répétées totalement méthylées (Hu et al. 2013). Ainsi, de nombreuses similitudes entre les profils de méthylation des mammifères et du poulet ont pu être observées.

La première analyse du méthylome de poulet à haute-résolution, permise grâce au WGBS, a été publiée en 2015 dans une étude portant sur la résistance à la maladie de Marek dans deux lignées de poulet (J. Li et al. 2015). Cette étude confirme la présence d'un niveau de méthylation plus important dans les exons que dans le reste du gène (Tian et al. 2013). Ces résultats mettent en évidence une corrélation entre la variation de la 5mC et une résistance à la maladie. En effet, l'étude de gènes liés à l'immunité, y compris le Toll Like Receptor 4 (TLR4), a montré des variations de méthylation et d'expression génique dans les deux lignées de poulets caractérisées pour avoir une résistance à la maladie de Marek différente (J. Li et al. 2015). Une autre étude WGBS récente a mis en évidence la variation des profils de 5mC en fonction de l'âge dans le tissu musculaire de poules pondeuses (M. Zhang et al. 2017). Les animaux âgés présentent un niveau de méthylation de l'ADN plus élevé que les jeunes individus. Cette augmentation de la méthylation a également été observée dans les zones promotrices de gènes impliqués dans la qualité de la viande (M. Zhang et al. 2017; Z. Zhang et al. 2018).

À ce jour, une seule analyse à l'échelle de l'épigénome a été effectuée dans une espèce aviaire différente du poulet, la mésange charbonnière (Laine et al. 2016). L'étude de la 5mC par WGBS dans le cerveau et le sang de cet oiseau a mis en évidence des taux réduits de méthylation sur les îlots CpG et autour des TSS. Par ailleurs, dans le cerveau, une corrélation négative entre les niveaux de méthylation dans le corps des gènes et leurs expressions a été observée (Laine et al. 2016). Enfin, le niveau de 5mC était plus important dans le cerveau que dans le sang, montrant à nouveau des différences dépendantes du tissu étudié (Derks et al. 2016).

5.3. Les modifications post-traductionnelles des histones chez le poulet

À ce jour, seuls quelques travaux ont étudié des MPTH chez les oiseaux (Luo et al. 2012; Mitra et al. 2012, 2015; Jahan et al. 2016). La première expérience concerne l'analyse des profils de MPTH dans la maladie de Marek dans deux lignées de poulet présentant des niveaux de résistance différents (Luo et al. 2012; Mitra et al. 2012, 2015). Ainsi, les profils de H3K4me3 et H3K27me3 ont été étudiés dans trois tissus impliqués dans la réponse immunitaire : la rate (Luo et al. 2012), le thymus (Mitra et al. 2012) et la bourse de Fabricius (Mitra et al. 2015). La marque H3K4me3 se retrouve principalement autour des TSS et est associée à des gènes actifs, comme précédemment montré chez les mammifères (Barski et al. 2007). Cela a également été confirmé récemment dans les érythrocytes de poulet (Jahan et al. 2016). La marque H3K27me3 est largement enrichie au niveau des TSS et dans le corps des gènes (Luo et al. 2012; Mitra et al. 2012, 2015) et associée à des gènes réprimés, comme observé dans d'autres espèces (Barski et al. 2007; David, Piégu, et al. 2017). De façon intéressante, des analyses comparatives ont révélé des régions différentiellement enrichies en ces MPTH entre les lignées de poulet sensibles et résistantes (Mitra et al. 2012, 2015), suggérant l'existence d'une « signature épigénétique » pour la sensibilité à la maladie. De plus, le profil de ces MPTH est altéré après l'infection au niveau des locus de gènes impliqués dans la réponse du système immunitaire (Luo et al. 2012). Par exemple, le niveau de H3K4me3 augmente sur le locus du gène CD8 α , de façon concomitante à l'expression de ce gène dans la rate (Luo et al. 2012). Dans le thymus, une tendance similaire a été observée au locus CTLA-4 (Mitra et al. 2012). Récemment, les profils de H3K27ac et H3K4me3 dans les érythrocytes nucléés ont été publiés chez le poulet (Jahan et al. 2016). Comme attendu, ces MPTH activatrices étaient associées à des gènes

transcriptionnellement actifs et leurs distributions à l'échelle du génome entier étaient similaires à celles rapportées dans d'autres espèces (Barski et al. 2007).

En conclusion, ces études tout-génome dans plusieurs lignées de poulet (Luo et al. 2012; Mitra et al. 2012, 2015; Jahan et al. 2016) montrent des profils de MPTH similaires entre les oiseaux et les mammifères, en accord avec la haute conservation des mécanismes épigénétiques entre espèces allant des plantes aux mammifères (Suzuki and Bird 2008).

Conclusion

Avec la démocratisation des techniques de SHD, de nombreuses avancées dans les connaissances des génomes aviaires et de leurs épigénomes ont été réalisées. Cependant, ces développements nécessitent de prendre en compte, dès la conception des projets, les diverses recommandations des consortiums internationaux et les besoins en analyses bioinformatiques et biostatistiques.

L'étude du méthylome et des profils des MPTH met en évidence de nombreuses caractéristiques communes entre l'épigénome du poulet et celui des mammifères : variations le long du génome liées à des caractéristiques génomiques telles que les TSS, corrélations de certaines marques présentes dans les régions promotrices avec l'expression des gènes, etc. Elles montrent des différences de profils de méthylation et de MPTH en fonction des tissus, de l'âge des individus ainsi que de la sensibilité des poulets aux pathogènes.

L'étude des épigénomes aviaires complétée par la cartographie des sites de chromatine ouverts (Buenrostro et al. 2015) ou l'analyse de la conformation du génome (Belton et al. 2012) devrait apporter une vision beaucoup plus claire des mécanismes de régulation de l'activité du génome liés au fonctionnement des organes, en contexte sain ou malade, ou plus généralement à la variation phénotypique. Ces données devraient permettre de mieux comprendre les mécanismes de réponse de l'épigénome face aux variations de l'environnement et donc faciliter la prise en compte progressive de ces mécanismes, quand c'est pertinent, dans l'amélioration des systèmes d'élevage, contribuant ainsi au développement d'une production avicole durable (Ibeagha-Awemu and Zhao 2015).

Remerciements

Ce travail est soutenu par la région Centre-Val de Loire, par le Département de génétique animale et le Département de physiologie animale et systèmes d'élevage (crédit incitatif INRA « Stressepimark ») de l'Institut National de la Recherche Agronomique (INRA). Il a été financé par l'ANR (programmes « ChickStress » - ANR-13-ADAP-0014 - et « QuailHeatE » - ANR-15-CE02-0009-01).

Références

- Andersson, Leif, Alan L Archibald, Cynthia D Bottema, Rudiger Brauning, Shane C Burgess, Dave W Burt, Eduardo Casas, et al. 2015. “Coordinated International Action to Accelerate Genome-to-Phenome with FAANG, the Functional Annotation of Animal Genomes Project.” *Genome Biology* 16 (March): 57. doi:10.1186/s13059-015-0622-4.
- Bae, Jae-Bum. 2013. “Perspectives of International Human Epigenome Consortium.” *Genomics & Informatics* 11 (1): 7. doi:10.5808/GI.2013.11.1.7.
- Baker, Monya. 2011. “Making Sense of Chromatin States.” *Nature Methods* 8 (9): 717–22. doi:10.1038/nmeth.1673.
- Bannister, Andrew J, and Tony Kouzarides. 2011. “Regulation of Chromatin by Histone Modifications.” *Cell Research* 21 (3). Nature Publishing Group: 381–95. doi:10.1038/cr.2011.22.
- Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. “High-Resolution Profiling of Histone Methylations in the Human Genome.” *Cell* 129 (4): 823–37. doi:10.1016/j.cell.2007.05.009.
- Belton, Jon-Matthew, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, and Job Dekker. 2012. “Hi-C: A Comprehensive Technique to Capture the Conformation of Genomes.” *Methods (San Diego, Calif.)* 58 (3): 268–76. doi:10.1016/j.ymeth.2012.05.001.
- Berger, Shelley L., Tony Kouzarides, Ramin Shiekhattar, and Ali Shilatifard. 2009. “An Operational Definition of Epigenetics.” *Genes & Development* 23 (7): 781–83. doi:10.1101/gad.1787609.
- Bernstein, Bradley E., John A. Stamatoyannopoulos, Joseph F. Costello, Bing Ren, Aleksandar Milosavljevic, Alexander Meissner, Manolis Kellis, et al. 2010. “The NIH Roadmap Epigenomics Mapping Consortium.” *Nature Biotechnology* 28 (10): 1045–48. doi:10.1038/nbt1010-1045.
- Bird, Adrian. 1992. “The Essentials of DNA Methylation.” *Cell* 70 (1): 5–8. doi:10.1016/0092-8674(92)90526-I.
- Bock, Christoph. 2012. “Analysing and Interpreting DNA Methylation Data.” *Nature Reviews. Genetics* 13 (10). Nature Publishing Group: 705–19. doi:10.1038/nrg3273.
- Buenrostro, Jason D., Beijing Wu, Howard Y. Chang, and William J. Greenleaf. 2015. “ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide.” *Current Protocols in Molecular Biology* 109 (January). Hoboken, NJ, USA: John Wiley & Sons, Inc.: 21.29.1-9. doi:10.1002/0471142727.mb2129s109.
- Busche, Stephan, Xiaojian Shao, Maxime Caron, Tony Kwan, Fiona Allum, Warren A. Cheung, Bing Ge, et al. 2015. “Population Whole-Genome Bisulfite Sequencing across Two Tissues Highlights the Environment as the Principal Source of Human Methylome Variation.” *Genome Biology* 16 (1). Genome Biology: 290. doi:10.1186/s13059-015-0856-1.
- Celniker, Susan E., Laura A. L. Dillon, Mark B. Gerstein, Kristin C. Gunsalus, Steven Henikoff, Gary H. Karpen, Manolis Kellis, et al. 2009. “Unlocking the Secrets of the Genome.” *Nature* 459 (7249): 927–30. doi:10.1038/459927a.
- Cokus, Shawn J, Suhua Feng, Xiaoyu Zhang, Zugen Chen, Barry Merriman, Christian D Haudenschild, Sriharsa Pradhan, Stanley F Nelson, Matteo Pellegrini, and Steven E Jacobsen. 2008. “Shotgun Bisulphite Sequencing of the Arabidopsis Genome Reveals DNA Methylation Patterning.” *Nature* 452 (7184): 215–19. doi:10.1038/nature06745.
- David, Sarah-Anne, Marjorie Mersch, Sylvain Foissac, Anne Collin, Frédérique Pitel, and Vincent Coustham. 2017. “Genome-Wide Epigenetic Studies in Chicken: A Review.” *Epigenomes* 1 (3): 20. doi:10.3390/epigenomes1030020.
- David, Sarah-Anne, Benoît Piégu, Christelle Hennequet-Antier, Maëlle Pannetier, Tiphaine Aguirre-Lavin, Sabine Crochet, Thierry Bordeau, et al. 2017. “An Assessment of Fixed and Native

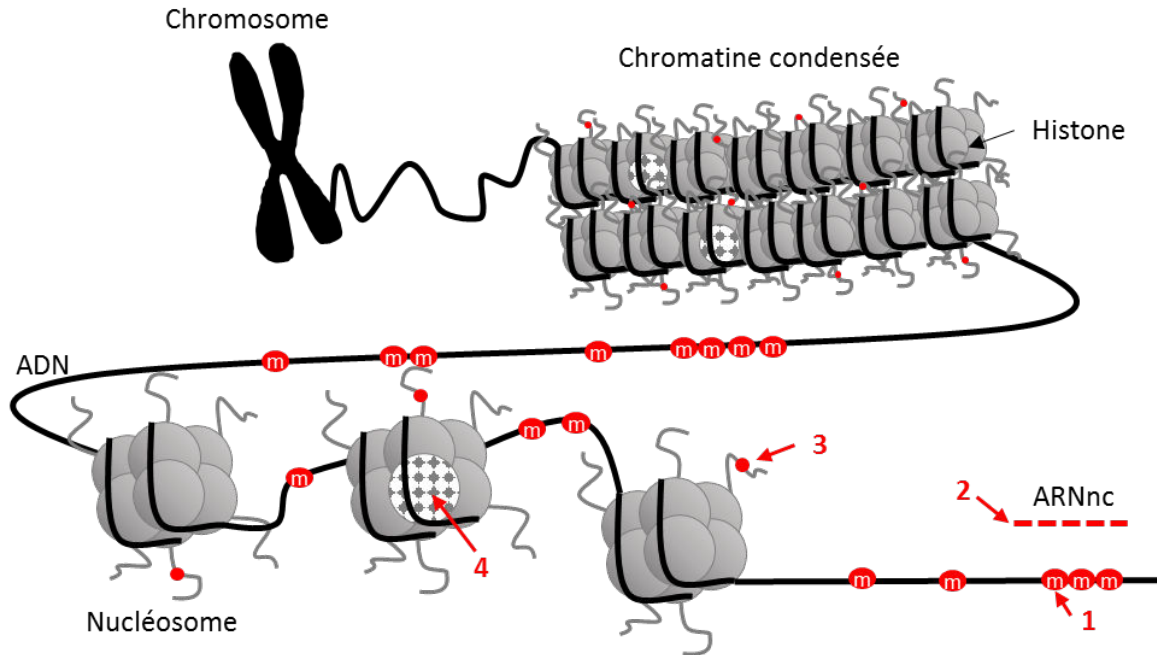
- Chromatin Preparation Methods to Study Histone Post-Translational Modifications at a Whole Genome Scale in Skeletal Muscle Tissue.” *Biological Procedures Online* 19 (1). Biological Procedures Online: 10. doi:10.1186/s12575-017-0059-0.
- Daxinger, Lucia, and Emma Whitelaw. 2010. “Transgenerational Epigenetic Inheritance: More Questions than Answers.” *Genome Research* 20 (12): 1623–28. doi:10.1101/gr.106138.110.
- Derks, Martijn F. L., Kyle M. Schachtschneider, Ole Madsen, Elio Schijlen, Koen J. F. Verhoeven, and Kees van Oers. 2016. “Gene and Transposable Element Methylation in Great Tit (*Parus Major*) Brain and Blood.” *BMC Genomics* 17 (1). BMC Genomics: 332. doi:10.1186/s12864-016-2653-y.
- Doherty, Rachael, and Christine Couldrey. 2014. “Exploring Genome Wide Bisulfite Sequencing for DNA Methylation Analysis in Livestock: A Technical Assessment.” *Frontiers in Genetics* 5 (May): 126. doi:10.3389/fgene.2014.00126.
- ENCODE Project Consortium. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489 (7414): 57–74. doi:10.1038/nature11247.
- Epigenome Blueprint. 2017. “ChIP-Seq Analysis Pipeline.” Accessed July 6. http://dcc.blueprint-epigenome.eu/#/md/chip_seq_grch37.
- Feil, Robert, and Mario F. Fraga. 2012. “Epigenetics and the Environment: Emerging Patterns and Implications.” *Nature Reviews. Genetics* 13 (2). Nature Publishing Group: 97–109. doi:10.1038/nrg3142.
- Frommer, Marianne, Louise E. McDonald, Douglas S. Millar, Christina M. Collis, Fujiko Watt, Geoffrey W. Grigg, Peter L Molloy, and Cheryl L. Paul. 1992. “A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands.” *Proceedings of the National Academy of Sciences of the United States of America* 89 (5): 1827–31. doi:10.1073/pnas.89.5.1827.
- Gardiner-Garden, M., and Marianne Frommer. 1987. “CpG Islands in Vertebrate Genomes.” *Journal of Molecular Biology* 196 (2): 261–82. doi:10.1016/0022-2836(87)90689-9.
- Globisch, Daniel, Martin Münzel, Markus Müller, Stylianos Michalakis, Mirko Wagner, Susanne Koch, Tobias Brückl, Martin Biel, and Thomas Carell. 2010. “Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates.” Edited by Anna Kristina Croft. *PLoS One* 5 (12): e15367. doi:10.1371/journal.pone.0015367.
- Gu, Hongcang, Zachary D. Smith, Christoph Bock, Patrick Boyle, Andreas Gnirke, and Alexander Meissner. 2011. “Preparation of Reduced Representation Bisulfite Sequencing Libraries for Genome-Scale DNA Methylation Profiling.” *Nature Protocols* 6 (4): 468–81. doi:10.1038/nprot.2010.190.
- Hahn, Maria A., Arthur X. Li, Xiwei Wu, and Gerd P. Pfeifer. 2015. “Single Base Resolution Analysis of 5-Methylcytosine and 5-Hydroxymethylcytosine by RRBS and TAB-RRBS.” Edited by Mukesh Verma. *Methods in Molecular Biology (Clifton, N.J.)*, Methods in Molecular Biology, 1238 (3). New York, NY: Springer New York: 273–87. doi:10.1007/978-1-4939-1804-1_14.
- Han, Yixing, and Ximiao He. 2016. “Integrating Epigenomics into the Understanding of Biomedical Insight.” *Bioinformatics and Biology Insights* 10: 267–89. doi:10.4137/BBI.S38427.
- Harris, R Alan, Ting Wang, Cristian Coarfa, Raman P. Nagarajan, Chibo Hong, Sara L. Downey, Brett E. Johnson, et al. 2010. “Comparison of Sequencing-Based Methods to Profile DNA Methylation and Identification of Monoallelic Epigenetic Modifications.” *Nature Biotechnology* 28 (10): 1097–1105. doi:10.1038/nbt.1682.
- Hu, Yongsheng, Haiping Xu, Zhenhui Li, Xuejuan Zheng, Xinzheng Jia, Qinghua Nie, and Xiquan Zhang. 2013. “Comparison of the Genome-Wide DNA Methylation Profiles between Fast-Growing and Slow-Growing Broilers.” Edited by Nicoletta Landsberger. *PLoS One* 8 (2):

- e56411. doi:10.1371/journal.pone.0056411.
- Ibeagha-Awemu, Eveline M., and Xin Zhao. 2015. “Epigenetic Marks: Regulators of Livestock Phenotypes and Conceivable Sources of Missing Variation in Livestock Improvement Programs.” *Frontiers in Genetics* 6 (SEP): 1–17. doi:10.3389/fgene.2015.00302.
- International Chicken Genome Sequencing Consortium. 2004. “Sequence and Comparative Analysis of the Chicken Genome Provide Unique Perspectives on Vertebrate Evolution.” *Nature* 432 (7018): 695–716. doi:10.1038/nature03154.
- Jahan, Sanzida, Wayne Xu, Shihua He, Carolina Gonzalez, Geneviève P. Delcuve, and James R. Davie. 2016. “The Chicken Erythrocyte Epigenome.” *Epigenetics & Chromatin* 9 (1). BioMed Central: 19. doi:10.1186/s13072-016-0068-2.
- Johnson, David S., Ali Mortazavi, Richard M. Myers, and Barbara Wold. 2007. “Genome-Wide Mapping of in Vivo Protein-DNA Interactions.” *Science (New York, N.Y.)* 316 (5830): 1497–1502. doi:10.1126/science.1141319.
- Jones, Peter A. 2012. “Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond.” *Nature Reviews. Genetics* 13 (7). Nature Publishing Group: 484–92. doi:10.1038/nrg3230.
- Jung, Youngsook L., Lovelace J. Luquette, Joshua W. K. Ho, Francesco Ferrari, Michael Tolstorukov, Aki Minoda, Robbyn Issner, et al. 2014. “Impact of Sequencing Depth in ChIP-Seq Experiments.” *Nucleic Acids Research* 42 (9): e74. doi:10.1093/nar/gku178.
- Kisliouk, Tatiana, and Noam Meiri. 2009. “A Critical Role for Dynamic Changes in Histone H3 Methylation at the Bdnf Promoter during Postnatal Thermotolerance Acquisition.” *European Journal of Neuroscience* 30 (December 2008): 1909–22. doi:10.1111/j.1460-9568.2009.06957.x.
- Kornberg, Roger D., and Yahli Lorch. 1999. “Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome.” *Cell* 98 (3): 285–94. doi:10.1016/S0092-8674(00)81958-3.
- Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705. doi:10.1016/j.cell.2007.02.005.
- Kurdyukov, Sergey, and Martyn Bullock. 2016. “DNA Methylation Analysis: Choosing the Right Method.” *Biology* 5 (1): 3. doi:10.3390/biology5010003.
- Laine, Veronika N., Toni I. Gossmann, Kyle M. Schachtschneider, Colin J. Garraway, Ole Madsen, Koen J. F. Verhoeven, Victor de Jager, et al. 2016. “Evolutionary Signals of Selection on Cognition from the Great Tit Genome and Methylome.” *Nature Communications* 7 (January): 10474. doi:10.1038/ncomms10474.
- Landt, Stephen G., Georgi K. Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E. Bernstein, et al. 2012. “ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia.” *Genome Research* 22 (9): 1813–31. doi:10.1101/gr.136184.111.
- Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3): R25. doi:10.1186/gb-2009-10-3-r25.
- Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics (Oxford, England)* 25 (14): 1754–60. doi:10.1093/bioinformatics/btp324.
- Li, Jinxiu, Rujiao Li, Ying Wang, Xiaoxiang Hu, Yiqiang Zhao, Li Li, Chungang Feng, et al. 2015. “Genome-Wide DNA Methylome Variation in Two Genetically Distinct Chicken Lines Using MethylC-Seq.” *BMC Genomics* 16 (1): 851. doi:10.1186/s12864-015-2098-8.
- Li, Qinghe, Ning Li, Xiaoxiang Hu, Jinxiu Li, Zhuo Du, Li Chen, Guangliang Yin, et al. 2011.

- “Genome-Wide Mapping of DNA Methylation in Chicken.” Edited by Pierre-Antoine Defossez. *PLoS One* 6 (5): e19428. doi:10.1371/journal.pone.0019428.
- Li, Ruiqiang, Chang Yu, Yingrui Li, Tak-Wah Lam, Siu-Ming Yiu, Karsten Kristiansen, and Jun Wang. 2009. “SOAP2: An Improved Ultrafast Tool for Short Read Alignment.” *Bioinformatics (Oxford, England)* 25 (15): 1966–67. doi:10.1093/bioinformatics/btp336.
- Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. 2009. “Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences.” *Nature* 462 (7271). Nature Publishing Group: 315–22. doi:10.1038/nature08514.
- Luo, Juan, Apratim Mitra, Fei Tian, Shuang Chang, Huanmin Zhang, Kairong Cui, Ying Yu, Keji Zhao, and Jiuzhou Song. 2012. “Histone Methylation Analysis and Pathway Predictions in Chickens after MDV Infection.” Edited by Huaijun Zhou. *PLoS One* 7 (7): e41849. doi:10.1371/journal.pone.0041849.
- Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461 (7265): 747–53. doi:10.1038/nature08494.
- Mardis, Elaine R. 2008. “The Impact of next-Generation Sequencing Technology on Genetics.” *Trends in Genetics : TIG* 24 (3): 133–41. doi:10.1016/j.tig.2007.12.007.
- McQueen, Heather A, Judy Fantes, Sally H Cross, Victoria H Clark, Alan L. Archibald, and Adrian P. Bird. 1996. “CpG Islands of Chicken Are Concentrated on Microchromosomes.” *Nature Genetics* 12 (3): 321–24. doi:10.1038/ng0396-321.
- Mitra, Apratim, Juan Luo, Yanghua He, Yulan Gu, Huanmin Zhang, Keji Zhao, Kairong Cui, and Jiuzhou Song. 2015. “Histone Modifications Induced by MDV Infection at Early Cytolytic and Latency Phases.” *BMC Genomics* 16 (1). ??? 311. doi:10.1186/s12864-015-1492-6.
- Mitra, Apratim, Juan Luo, Huanming Zhang, Kairong Cui, Keji Zhao, and Jiuzhou Song. 2012. “Marek’s Disease Virus Infection Induces Widespread Differential Chromatin Marks in Inbred Chicken Lines.” *BMC Genomics* 13 (1): 557. doi:10.1186/1471-2164-13-557.
- Ponnaluri, V. K. Chaithanya, Kenneth C Ehrlich, Guoqiang Zhang, Michelle Lacey, Douglas Johnston, Sriharsa Pradhan, and Melanie Ehrlich. 2017. “Association of 5-Hydroxymethylation and 5-Methylation of DNA Cytosine with Tissue-Specific Gene Expression.” *Epigenetics* 12 (2): 123–38. doi:10.1080/15592294.2016.1265713.
- Rhee, Ho Sung, and B. Franklin Pugh. 2011. “Comprehensive Genome-Wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution.” *Cell* 147 (6). Elsevier Inc.: 1408–19. doi:10.1016/j.cell.2011.11.013.
- Rigai, Guillem, Sandrine Balzergue, Véronique Brunaud, Eddy Blondet, Andrea Rau, Odile Rogier, José Caius, et al. 2016. “Synthetic Data Sets for the Identification of Key Ingredients for RNA-Seq Differential Analysis.” *Briefings in Bioinformatics*, no. August (October): bbw092. doi:10.1093/bib/bbw092.
- Robertson, Gordon, Martin Hirst, Matthew Bainbridge, Misha Bilenky, Yongjun Zhao, Thomas Zeng, Ghia Euskirchen, et al. 2007. “Genome-Wide Profiles of STAT1 DNA Association Using Chromatin Immunoprecipitation and Massively Parallel Sequencing.” *Nature Methods* 4 (8): 651–57. doi:10.1038/nmeth1068.
- Roundtree, Ian A., Molly E. Evans, Tao Pan, and Chuan He. 2017. “Dynamic RNA Modifications in Gene Expression Regulation.” *Cell* 169 (7): 1187–1200. doi:10.1016/j.cell.2017.05.045.
- Simpson, Jared T., Rachael E. Workman, P. C. Zuzarte, Matei David, L. J. Dursi, and Winston Timp. 2017. “Detecting DNA Cytosine Methylation Using Nanopore Sequencing.” *Nature Methods* 14 (4): 407–10. doi:10.1038/nmeth.4184.
- Smith, Zachary D., Hongcang Gu, Christoph Bock, Andreas Gnirke, and Alexander Meissner. 2009.

- “High-Throughput Bisulfite Sequencing in Mammalian Genomes.” *Methods (San Diego, Calif.)* 48 (3). Elsevier Inc.: 226–32. doi:10.1016/j.ymeth.2009.05.003.
- Stein, Lincoln. 2001. “Genome Annotation: From Sequence to Biology.” *Nature Reviews. Genetics* 2 (7): 493–503. doi:10.1038/35080529.
- Suzuki, Miho M., and Adrian Bird. 2008. “DNA Methylation Landscapes: Provocative Insights from Epigenomics.” *Nature Reviews. Genetics* 9 (6): 465–76. doi:10.1038/nrg2341.
- Tessarz, Peter, and Tony Kouzarides. 2014. “Histone Core Modifications Regulating Nucleosome Structure and Dynamics.” *Nature Reviews. Molecular Cell Biology* 15 (11). Nature Publishing Group: 703–8. doi:10.1038/nrm3890.
- The ENCODE Project Consortium. 2017. “Whole-Genome Bisulfite Sequencing Data Standards and Processing Pipeline.” Accessed July 6. <https://www.encodeproject.org/wgbs/>.
- Tian, Fei, Fei Zhan, Nathan D. VanderKraats, Jeffrey F. Hiken, John R. Edwards, Huanmin Zhang, Keji Zhao, and Jiuzhou Song. 2013. “DNMT Gene Expression and Methylome in Marek’s Disease Resistant and Susceptible Chickens prior to and Following Infection by MDV.” *Epigenetics* 8 (4): 431–44. doi:10.4161/epi.24361.
- Triantaphyllopoulos, Kostas A., Ioannis Ikonopoulos, and Andrew J. Bannister. 2016. “Epigenetics and Inheritance of Phenotype Variation in Livestock.” *Epigenetics & Chromatin* 9 (1). BioMed Central: 31. doi:10.1186/s13072-016-0081-5.
- Urich, Mark A., Joseph R. Nery, Ryan Lister, Robert J. Schmitz, and Joseph R. Ecker. 2015. “MethylC-Seq Library Preparation for Base-Resolution Whole-Genome Bisulfite Sequencing.” *Nature Protocols* 10 (3): 475–83. doi:10.1038/nprot.2014.114.
- Verhulst, Eveline C., A. Christa Mateman, Mathijs V. Zwier, Samuel P. Caro, Koen J.F. Verhoeven, and Kees van Oers. 2016. “Evidence from Pyrosequencing Indicates That Natural Variation in Animal Personality Is Associated with DRD4 DNA Methylation.” *Molecular Ecology* 25 (8): 1801–11. doi:10.1111/mec.13519.
- Wolff, G L, R L Kodell, S R Moore, and C A Cooney. 1998. “Maternal Epigenetics and Methyl Supplements Affect Agouti Gene Expression in Avy/a Mice.” *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 12 (11): 949–57. <http://www.ncbi.nlm.nih.gov/pubmed/9707167>.
- Zhang, Meng, Feng-Bin Yan, Fang Li, Ke-Ren Jiang, Dong-Hua Li, Rui-Li Han, Zhuan-Jan Li, et al. 2017. “Genome-Wide DNA Methylation Profiles Reveal Novel Candidate Genes Associated with Meat Quality at Different Age Stages in Hens.” *Scientific Reports* 7 (April): 45564. doi:10.1038/srep45564.
- Zhang, Tianyi, Sarah Cooper, and Neil Brockdorff. 2015. “The Interplay of Histone Modifications - Writers That Read.” *EMBO Reports* 16 (11): 1467–81. doi:10.15252/embr.201540945.
- Zhang, Zengrong, Huarui Du, Lijun Bai, Chaowu Yang, Qingyun Li, Xiaocheng Li, Mohan Qiu, et al. 2018. “Whole Genome Bisulfite Sequencing Reveals Unique Adaptations to High-Altitude Environments in Tibetan Chickens.” *PLoS ONE* 13 (3). doi:10.1371/journal.pone.0193597.
- Ziller, Michael J., Kasper D. Hansen, Alexander Meissner, and Martin J. Aryee. 2015. “Coverage Recommendations for Methylation Analysis by Whole-Genome Bisulfite Sequencing.” *Nature Methods* 12 (3): 230–2, 1 p following 232. doi:10.1038/nmeth.3152.

Illustrations



Les principales modifications épigénétiques :

1 – Méthylation de l'ADN

2 – ARN non codants (ARNnc)

3 – Modification post-traductionnelle des histones (MPTH)

4 – Variant d'histones

Figure 1. Schéma des principales modifications épigénétiques.

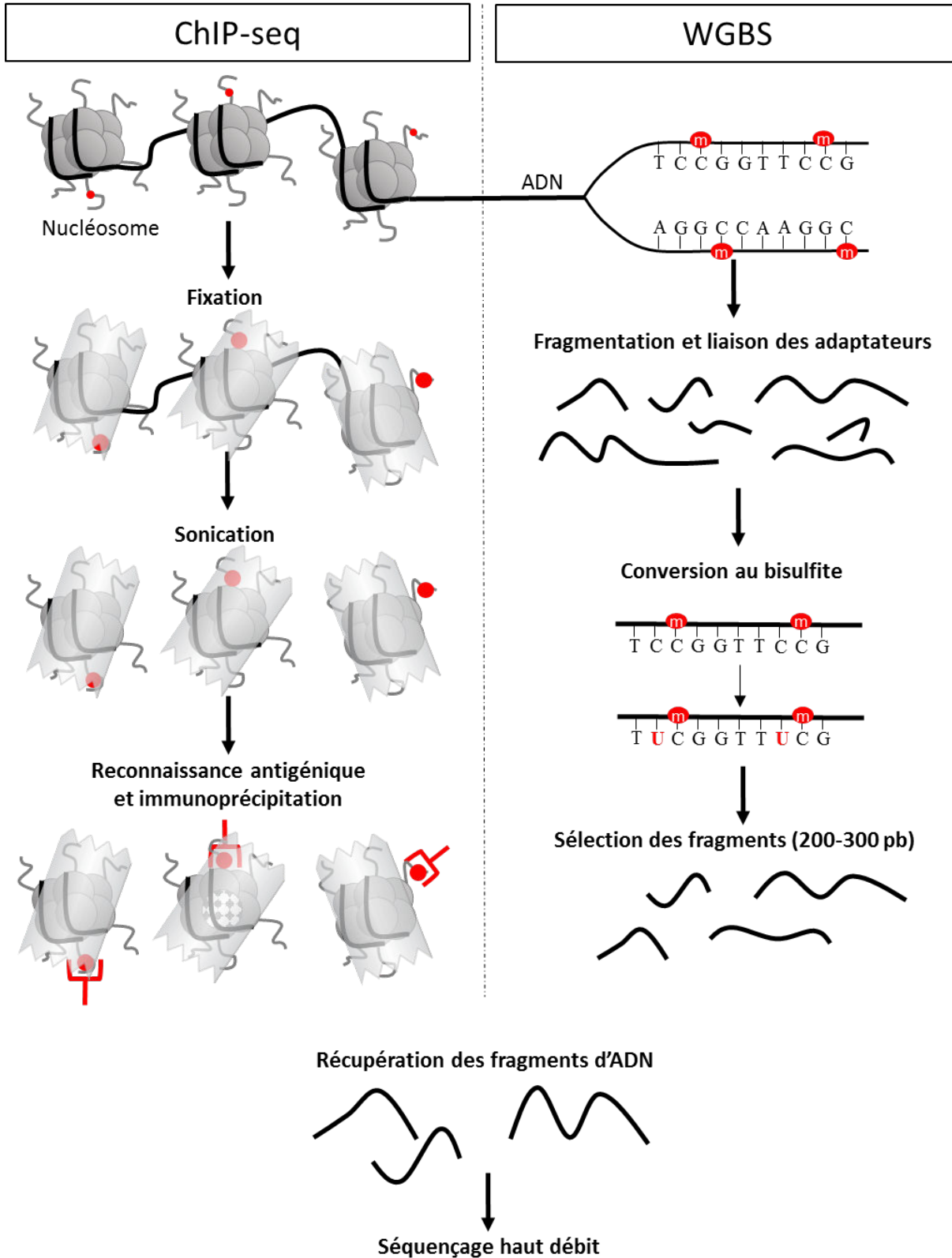


Figure 2. Représentation graphique des protocoles d'immunoprécipitation de la chromatine (ChIP-seq) et de *Whole Genome Bisulfite Sequencing* (WGBS).

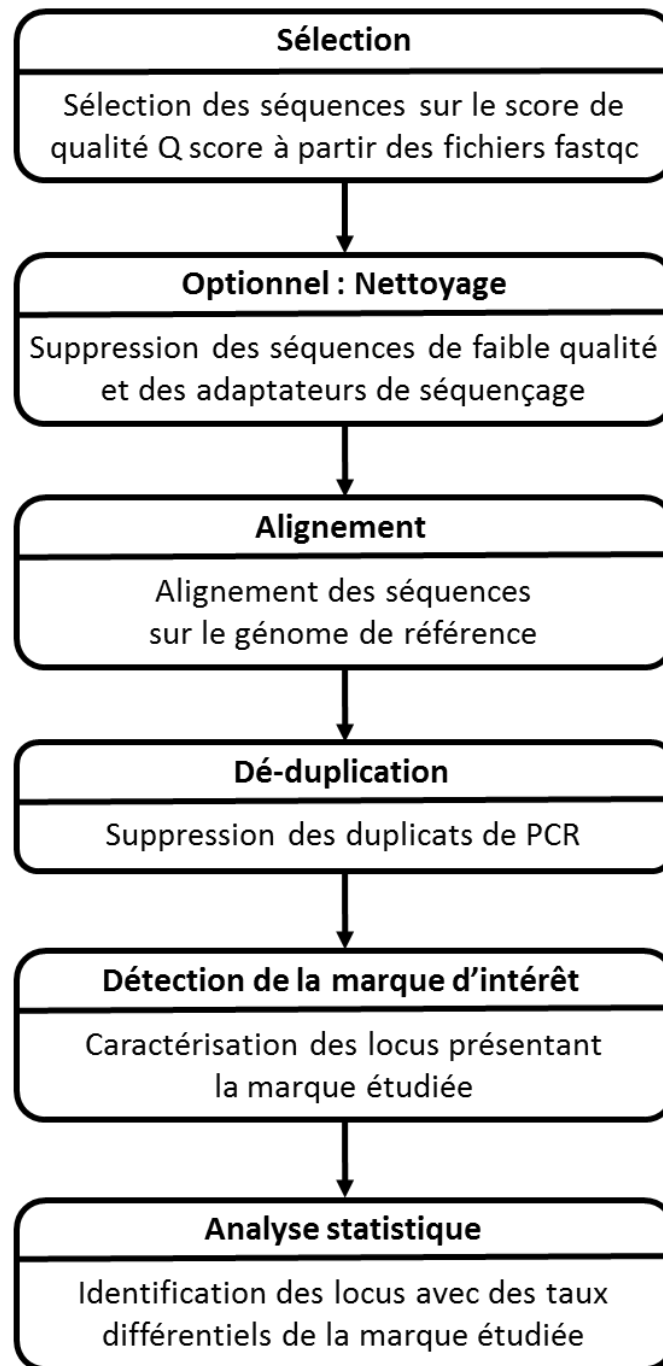


Figure 3. Exemple d'analyse bioinformatique de données d'immunoprécipitation de la chromatine (ChIP-seq) et de *Whole Genome Bisulfite Sequencing* (WGBS).

Modification épigénétique	Nom de la technique	Préparation de l'ADN	Principe d'identification	Fraction ADN étudiée	Résolution
Méthylation de l'ADN	<i>Methylation sensitive Restriction Enzyme sequencing (MRE-seq)</i>	Aucune	Enzyme de restriction sensible à la méthylation + Séquençage haut-débit	Génome entier	Dépend des sites de coupure enzymatique
	<i>Methylated DNA ImmunoPrecipitation (MeDIP-seq ou MBD-seq)</i>	Aucune	Reconnaissance par anticorps de l'ADN méthylé + Séquençage haut-débit	Génome entier	Dépend de la taille des fragments immunoprécipités
	<i>Whole Genome Bisulfite Sequencing (WGBS ou MethylC-seq)</i>	Conversion au bisulfite de sodium	Séquençage haut-débit	Génome entier	Base
	<i>Reduced representation bisulfite sequencing (RRBS)</i>	Conversion au bisulfite de sodium	Enzyme de restriction + Séquençage haut-débit	Sous-représentation du génome, en général enrichie en îlots CpG	Base
Modification post-traductionnelle des histones	Immunoprécipitation de la chromatine (ChIP-seq)	Aucune	Reconnaissance par anticorps + Séquençage haut-débit	Génome entier	Mono/di-nucléosome

Tableau 1. Liste non exhaustive de technologies d'analyse de la méthylation de l'ADN et des modifications post-traductionnelles des histones.

Bibliographie

- [1] IPCC, Summary for Policymakers, book section SPM, pp. 1–30. Cambridge, United Kingdom and New York, NY, USA : Cambridge University Press, 2013.
- [2] IPCC, Summary for Policymakers. Cambridge, United Kingdom and New York, NY, USA : Cambridge University Press, 2007.
- [3] P. J. Gerber, H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio, Tackling Climate Change through Livestock - A global assessment of emissions and mitigation opportunities, 2013.
- [4] G. Amand, C. Aubert, C. Bourdette, I. Bouvarel, D. Chevalier, A. Dusantier, Y. Franck, M. Guillou, M. Hassouna, R. Le Biavan, F. Mahe, J.-P. Prigent, and P. Robin, “La prévention du coup de chaleur en aviculture,” La Revue Scientifique de l’Aviculture, vol. Hors série, pp. 1–66, 2004.
- [5] A. Bordas and P. Merat, “Effects of the Naked-Neck Gene on Traits Associated with Egg Laying in a Dwarf Stock at Two Temperatures,” British Poultry Science, vol. 25, no. 2, pp. 195–207, 1984.
- [6] A. Bordas and F. Minvielle, “Réponse à la chaleur de poules pondeuses issues de lignées sélectionnées pour une faible (R-) ou forte (R +) consommation alimentaire résiduelle,” Genetics Selection Evolution, vol. 29, no. 3, pp. 279–290, 1997.
- [7] D. F. Pereira and I. A. Nääs, “Estimating the thermoneutral zone for broiler breeders using behavioral analysis,” Computers and Electronics in Agriculture, vol. 62, no. 1, pp. 2–7, 2008.
- [8] S. Mutaf, N. Seber Kahraman, and M. Z. Firat, “Intermittent partial surface wetting and its effect on body-surface temperatures and egg production of white and brown domestic laying hens in Antalya (Turkey),” British Poultry Science, vol. 50, no. 1, pp. 33–38, 2009.
- [9] L. J. Lara and M. H. Rostagno, “Impact of heat stress on poultry production,” Animals, vol. 3, no. 2, pp. 356–369, 2013.
- [10] S. Jiang, L. Cui, C. Shi, X. Ke, J. Luo, and J. Hou, “Effects of dietary energy and calcium levels on performance, egg shell quality and bone metabolism in hens,” Veterinary Journal, vol. 198, no. 1, pp. 252–258, 2013.
- [11] W. M. Quinteiro-Filho, M. V. Rodrigues, A. Ribeiro, V. Ferraz-de Paula, M. L. Pinheiro, L. R. M. Sá, A. J. P. Ferreira, and J. Palermo-Neto, “Acute heat stress impairs performance parameters and induces mild intestinal enteritis in broiler chickens : Role of acute hypothalamic-pituitary-adrenal axis activation1,” Journal of Animal Science, vol. 90, pp. 1986–1994, jun 2012.
- [12] X. Tao, Z. Y. Zhang, H. Dong, H. Zhang, and H. Xin, “Responses of thyroid hormones of market-size broilers to thermoneutral constant and warm cyclic temperatures,” Poultry science, vol. 85, no. 9, pp. 1520–8, 2006.
- [13] S. Laisse, R. Baumont, T. Turini, L. Dusart, D. Gaudré, B. Rouillé, M. Benoit, P.-M. Rosner, and J.-L. Peyraud, “Efficience alimentaire des élevages : un nouveau regard sur la compétition entre alimentation animale et humaine,” Colloque du GIS Elevages Demain, pp. 1–12, 2017.
- [14] X. Wang, H. Zhang, H. Wang, J. Wang, S. Wu, and G. Qi, “Effect of dietary protein sources on production performance, egg quality, and plasma parameters of laying hens,” Asian-Australasian Journal of Animal Sciences, vol. 30, pp. 400–409, sep 2016.
- [15] E. Valkonen, E. Venalainen, L. Rossow, and J. Valaja, “Effects of Dietary Energy Content on the Performance of Laying Hens in Furnished and Conventional Cages,” Poultry Science, vol. 87, pp. 844–852, may 2008.

- [16] L. Lacassagne, "Alimentation des volailles : substituts au tourteau de soja," INRA Production Animales, vol. 1, no. 1, pp. 47–57, 1988.
- [17] L. Lacassagne, "Alimentation des volailles : substituts au tourteau de soja," INRA Production Animales, vol. 1, no. 2, pp. 123–128, 1988.
- [18] A. Pérez-Bonilla, S. Novoa, J. García, M. Mohiti-Asli, M. Frikha, and G. G. Mateos, "Effects of energy concentration of the diet on productive performance and egg quality of brown egg-laying hens differing in initial body weight," Poultry Science, vol. 91, no. 12, pp. 3156–3166, 2012.
- [19] R. M. Koch, L. A. Swiger, D. Chambers, and K. E. Gregory, "Efficiency of Feed Use in Beef Cattle," Journal of Animal Science, vol. 22, pp. 486–494, may 1963.
- [20] A. Bordas, P. Mérat, G. Coquerelle, and J. P. Noé, "Influence d'un aliment dilué sur des lignées de poules pondeuses sélectionnées sur la consommation alimentaire résiduelle," Genetics Selection Evolution, vol. 27, no. 3, p. 299, 1995.
- [21] E. M. Goldberg, D. Ryland, R. A. Gibson, M. Aliani, and J. D. House, "Designer laying hen diets to improve egg fatty acid profile and maintain sensory quality," Food Science & Nutrition, vol. 1, pp. 324–335, jul 2013.
- [22] G. G. Mateos, E. Jimenez-Moreno, M. P. Serrano, and R. P. Lazaro, "Poultry response to high levels of dietary fiber sources varying in physical and chemical characteristics," The Journal of Applied Poultry Research, vol. 21, pp. 156–174, mar 2012.
- [23] W. M. Rauw and L. Gomez-Raya, "Genotype by environment interaction and breeding for robustness in livestock," Frontiers in Genetics, vol. 6, pp. 1–15, oct 2015.
- [24] A. Van Goor, K. J. Bolek, C. M. Ashwell, M. E. Persia, M. F. Rothschild, C. J. Schmidt, and S. J. Lamont, "Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress," Genetics Selection Evolution, vol. 47, p. 96, dec 2015.
- [25] R. Feil and M. F. Fraga, "Epigenetics and the environment : emerging patterns and implications," Nature Reviews Genetics, vol. 13, no. 2, pp. 97–109, 2012.
- [26] W.-S. Yong, F.-M. Hsu, and P.-Y. Chen, "Profiling genome-wide DNA methylation," Epigenetics & Chromatin, vol. 9, p. 26, dec 2016.
- [27] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," Cell Research, vol. 21, pp. 381–395, mar 2011.
- [28] X. Li, X. Xiong, and C. Yi, "Epitranscriptome sequencing technologies : decoding RNA modifications," Nature Methods, vol. 14, pp. 23–31, jan 2017.
- [29] C. D. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control," Nature Reviews Genetics, vol. 17, pp. 487–500, jun 2016.
- [30] E. M. Ibeagha-Awemu and X. Zhao, "Epigenetic marks : regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs," Frontiers in Genetics, vol. 6, pp. 1–17, sep 2015.
- [31] O. González-Recio, M. A. Toro, and A. Bach, "Past, present, and future of epigenetics applied to livestock breeding," Frontiers in Genetics, vol. 6, pp. 1–5, sep 2015.
- [32] K. A. Triantaphyllopoulos, I. Ikonopoulou, and A. J. Bannister, "Epigenetics and inheritance of phenotype variation in livestock," Epigenetics & chromatin, vol. 9, p. 31, dec 2016.
- [33] T. Kouzarides, "Chromatin Modifications and Their Function," Cell, vol. 128, no. 4, pp. 693–705, 2007.
- [34] H. Dai and Z. Wang, "Histone Modification Patterns and Their Responses to Environment," Current Environmental Health Reports, vol. 1, pp. 11–21, mar 2014.
- [35] A. A. Cronican, N. F. Fitz, A. Carter, M. Saleem, S. Shiva, A. Barchowsky, R. Koldamova, J. Schug, and I. Lefterov, "Genome-Wide Alteration of Histone H3K9 Acetylation Pattern in Mouse Offspring Prenatally Exposed to Arsenic," PLoS ONE, vol. 8, p. e53478, feb 2013.
- [36] V. Coustham, P. Li, A. Strange, C. Lister, J. Song, and C. Dean, "Quantitative Modulation of Polycomb Silencing Underlies Natural Variation in Vernalization," Science, vol. 337, pp. 584–587, aug 2012.

- [37] Y. Zhao and B. A. Garcia, "Comprehensive Catalog of Currently Documented Histone Modifications," Cold Spring Harbor Perspectives in Biology, vol. 7, p. a025064, sep 2015.
- [38] C. S. Technology, "Histone Modification Table," 2018.
- [39] E. Bernstein and C. D. Allis, "RNA meets chromatin," Genes & Development, vol. 19, pp. 1635–1655, jul 2005.
- [40] J. S. Mattick, "Non-coding RNAs : the architects of eukaryotic complexity," EMBO reports, vol. 2, pp. 986–991, nov 2001.
- [41] J. Bao, J. Wu, A. S. Schuster, G. W. Hennig, and W. Yan, "Expression Profiling Reveals Developmentally Regulated lncRNA Repertoire in the Mouse Male Germline1," Biology of Reproduction, vol. 89, pp. 1–12, nov 2013.
- [42] J. Wang, X. Meng, O. B. Dobrovolskaya, Y. L. Orlov, and M. Chen, "Non-coding RNAs and Their Roles in Stress Response in Plants," Genomics, Proteomics & Bioinformatics, vol. 15, pp. 301–312, oct 2017.
- [43] P. P. Amaral, M. E. Dinger, and J. S. Mattick, "Non-coding RNAs in homeostasis, disease and stress responses : an evolutionary perspective," Briefings in Functional Genomics, vol. 12, pp. 254–278, may 2013.
- [44] M. B. Clark, R. L. Johnston, M. Inostroza-Ponta, A. H. Fox, E. Fortini, P. Moscato, M. E. Dinger, and J. S. Mattick, "Genome-wide analysis of long noncoding RNA stability," Genome Research, vol. 22, pp. 885–898, may 2012.
- [45] J. Krol, V. Busskamp, I. Markiewicz, M. B. Stadler, S. Ribí, J. Richter, J. Duebel, S. Bicker, H. J. Fehling, D. Schübeler, T. G. Oertner, G. Schratt, M. Bibel, B. Roska, and W. Filipowicz, "Characterizing Light-Regulated Retinal MicroRNAs Reveals Rapid Turnover as a Common Property of Neuronal MicroRNAs," Cell, vol. 141, pp. 618–631, may 2010.
- [46] P. P. Amaral and J. S. Mattick, "Noncoding RNA in development," Mammalian Genome, vol. 19, pp. 454–492, aug 2008.
- [47] S. P. Hazen, F. Naef, T. Quisel, J. M. Gendron, H. Chen, J. R. Ecker, J. O. Borevitz, and S. A. Kay, "Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays," Genome Biology, vol. 10, no. 2, p. R17, 2009.
- [48] S. L. Coon, P. J. Munson, P. F. Cherukuri, D. Sugden, M. F. Rath, M. Moller, S. J. H. Clokie, C. Fu, M. E. Olanich, Z. Rangel, T. Werner, J. C. Mullikin, D. C. Klein, B. Benjamin, R. Blakesley, G. Bouffard, S. Brooks, G. Chu, H. Coleman, M. Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove, S.-l. Ho, T. Johnson, R. Legaspi, S. Lovett, Q. Maduro, C. Masiello, B. Maskeri, J. McDowell, C. Montemayor, B. Novotny, M. Park, N. Riebow, K. Schandler, B. Schmidt, C. Sison, M. Stantripop, J. Thomas, M. Vemulapalli, and A. Young, "Circadian changes in long noncoding RNAs in the pineal gland," Proceedings of the National Academy of Sciences, vol. 109, pp. 13319–13324, aug 2012.
- [49] W. Yu, D. Gius, P. Onyango, K. Muldoon-Jacobs, J. Karp, A. P. Feinberg, and H. Cui, "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA," Nature, vol. 451, pp. 202–206, jan 2008.
- [50] S. C. Lakhotia, "Long non-coding RNAs coordinate cellular responses to stress," Wiley Interdisciplinary Reviews : RNA, vol. 3, pp. 779–796, nov 2012.
- [51] S. Sengupta and S. C. Lakhotia, "Altered expressions of the noncoding hsr-omega gene enhances poly-Q-induced neurotoxicity in Drosophila.," RNA biology, vol. 3, no. 1, pp. 28–35, 2006.
- [52] K. V. Prasanth, T. K. Rajendra, A. K. Lal, and S. C. Lakhotia, "Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila.," Journal of cell science, vol. 113 Pt 19, pp. 3485–97, oct 2000.
- [53] R. Arya, M. Mallik, and S. C. Lakhotia, "Heat shock genes — integrating cell survival and death," Journal of Biosciences, vol. 32, pp. 595–610, apr 2007.
- [54] W. Bendena, A. Ayme-Southgate, J. Garbe, and M. Pardue, "Expression of heat-shock locus hsr-omega in nonstressed cells during development in Drosophila melanogaster," Developmental Biology, vol. 144, pp. 65–77, mar 1991.
- [55] T. K. Rajendra, K. V. Prasanth, and S. C. Lakhotia, "Male sterility associated with overexpression of the noncoding hsr Ω gene in cyst cells of testis of Drosophila melanogaster," Journal of Genetics, vol. 80, pp. 97–110, aug 2001.

- [56] M. G. Tapadia and S. C. Lakhotia, "Specific induction of the hsr omega locus of *Drosophila melanogaster* by amides.," Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, vol. 5, pp. 359–62, sep 1997.
- [57] T. A. Allen, S. Von Kaenel, J. A. Goodrich, and J. F. Kugel, "The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock," Nature Structural & Molecular Biology, vol. 11, pp. 816–821, sep 2004.
- [58] P. Yakovchuk, J. A. Goodrich, and J. F. Kugel, "B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes," Proceedings of the National Academy of Sciences, vol. 106, pp. 5569–5574, apr 2009.
- [59] C. Jolly and S. C. Lakhotia, "Human sat III and *Drosophila* hsr ω transcripts : a common paradigm for regulation of nuclear RNA processing in stressed cells," Nucleic Acids Research, vol. 34, pp. 5508–5514, nov 2006.
- [60] R. Valgardsdottir, I. Chiodi, M. Giordano, A. Rossi, S. Bazzini, C. Ghigna, S. Riva, and G. Biamonti, "Transcription of Satellite III non-coding RNAs is a general stress response in human cells," Nucleic Acids Research, vol. 36, pp. 423–434, feb 2008.
- [61] I. Shamovsky and E. Nudler, "Gene Control by Large Noncoding RNAs," Science's STKE, vol. 2006, pp. pe40–pe40, sep 2006.
- [62] X. Wang, Z. Gu, and H. Jiang, "MicroRNAs in farm animals," animal, vol. 7, pp. 1567–1575, oct 2013.
- [63] J. M. Romao, W. Jin, M. He, T. McAllister, and L. L. Guan, "Altered microRNA expression in bovine subcutaneous and visceral adipose tissues from cattle under different diet," PLoS ONE, vol. 7, no. 7, 2012.
- [64] Y. Saletore, K. Meyer, J. Korlach, I. D. Vilfan, S. Jaffrey, and C. E. Mason, "The birth of the Epitranscriptome : deciphering the function of RNA modifications," Genome Biology, vol. 13, no. 10, p. 175, 2012.
- [65] D. Holoch and D. Moazed, "RNA-mediated epigenetic regulation of gene expression," Nature Reviews Genetics, vol. 16, pp. 71–84, feb 2015.
- [66] N. Liu and T. Pan, "RNA epigenetics," Translational Research, vol. 165, pp. 28–35, jan 2015.
- [67] L. Scourzic, E. Mouly, and O. A. Bernard, "TET proteins and the control of cytosine demethylation in cancer," Genome Medicine, vol. 7, no. 1, p. 9, 2015.
- [68] T. Ordog, S. A. Syed, Y. Hayashi, and D. T. Asuzu, "Epigenetics and chromatin dynamics : a review and a paradigm for functional disorders," Neurogastroenterology & Motility, vol. 24, pp. 1054–1068, dec 2012.
- [69] A. C. Mirabella, B. M. Foster, and T. Bartke, "Chromatin deregulation in disease," Chromosoma, vol. 125, pp. 75–93, mar 2016.
- [70] L. Ho and G. R. Crabtree, "Chromatin remodelling during development," Nature, vol. 463, pp. 474–484, jan 2010.
- [71] J. Lämke and I. Bäurle, "Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants," Genome Biology, vol. 18, p. 124, dec 2017.
- [72] V. Chinnusamy, Z. Gong, and J.-K. Zhu, "Abscisic Acid-mediated Epigenetic Processes in Plant Development and Stress Responses," Journal of Integrative Plant Biology, vol. 50, pp. 1187–1195, oct 2008.
- [73] S.-K. Han, Y. Sang, A. Rodrigues, M.-F. Wu, P. L. Rodriguez, and D. Wagner, "The SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA Represses Abscisic Acid Responses in the Absence of the Stress Stimulus in *Arabidopsis*," The Plant Cell, vol. 24, pp. 4892–4906, dec 2012.
- [74] E. A. Mazzi and K. F. Soliman, "Basic concepts of epigenetics : impact of environmental signals on gene expression," Epigenetics, vol. 7, pp. 119–130, feb 2012.
- [75] F. Lyko, "The DNA methyltransferase family : A versatile toolkit for epigenetic regulation," Nature Reviews Genetics, vol. 19, no. 2, pp. 81–92, 2018.
- [76] E. Li and Y. Zhang, "DNA Methylation in Mammals," Cold Spring Harbor Perspectives in Biology, vol. 6, pp. a019133–a019133, may 2014.

- [77] M. Tahiliani, K. P. Koh, Y. Shen, W. a. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao, "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1," *Science*, vol. 324, pp. 930–935, may 2009.
- [78] S. Kriaucionis and N. Heintz, "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain," *Science*, vol. 324, pp. 929–930, may 2009.
- [79] L. Shen, C.-X. Song, C. He, and Y. Zhang, "Mechanism and Function of Oxidative Reversal of DNA and RNA Methylation," *Annual Review of Biochemistry*, vol. 83, pp. 585–614, jun 2014.
- [80] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang, "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine," *Science*, vol. 333, pp. 1300–1303, sep 2011.
- [81] X. Lu, D. Han, B. S. Zhao, C.-X. Song, L.-S. Zhang, L. C. Doré, and C. He, "Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics," *Cell Research*, vol. 25, pp. 386–389, mar 2015.
- [82] G.-Z. Luo, M. A. Blanco, E. L. Greer, C. He, and Y. Shi, "DNA N6-methyladenine : a new epigenetic mark in eukaryotes?," *Nature Reviews Molecular Cell Biology*, vol. 16, pp. 705–710, dec 2015.
- [83] T. P. Wu, T. Wang, M. G. Seetin, Y. Lai, S. Zhu, K. Lin, Y. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. Wang, L. S. Hon, G. Fang, J. A. Swenberg, and A. Z. Xiao, "DNA methylation on N6-adenine in mammalian embryonic stem cells," *Nature*, vol. 532, pp. 329–333, mar 2016.
- [84] M. A. Sánchez-Romero, I. Cota, and J. Casadesús, "DNA methylation in bacteria : from the methyl group to the methylome," *Current Opinion in Microbiology*, vol. 25, pp. 9–16, jun 2015.
- [85] M. M. Suzuki and A. P. Bird, "DNA methylation landscapes : provocative insights from epigenomics," *Nature Reviews Genetics*, vol. 9, pp. 465–476, jun 2008.
- [86] A. Zemach, I. E. McDaniel, P. Silva, and D. Zilberman, "Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation," *Science*, vol. 328, pp. 916–919, may 2010.
- [87] S. Feng, S. J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini, and S. E. Jacobsen, "Conservation and divergence of methylation patterning in plants and animals," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 8689–8694, may 2010.
- [88] J. A. Yoder, C. P. Walsh, and T. H. Bestor, "Cytosine methylation and the ecology of intragenomic parasites," *Trends in Genetics*, vol. 13, pp. 335–340, aug 1997.
- [89] E. U. Selker, "Epigenetic phenomena in filamentous fungi : useful paradigms or repeat-induced confusion?," *Trends in Genetics*, vol. 13, pp. 296–301, aug 1997.
- [90] Z. Lippman, A.-V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia, W. Richard McCombie, K. Lavine, V. Mittal, B. May, K. D. Kasschau, J. C. Carrington, R. W. Doerge, V. Colot, and R. Martienssen, "Role of transposable elements in heterochromatin and epigenetic control," *Nature*, vol. 430, pp. 471–476, jul 2004.
- [91] J. A. Law and S. E. Jacobsen, "Establishing, maintaining and modifying DNA methylation patterns in plants and animals," *Nature Reviews Genetics*, vol. 11, pp. 204–220, mar 2010.
- [92] P. Neddermann, P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J. J. Hsuan, K. Wiebauer, and J. Jiricny, "Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase.," *The Journal of biological chemistry*, vol. 271, pp. 12767–74, may 1996.
- [93] A. Maiti, M. S. Noon, A. D. MacKerell, E. Pozharski, and A. C. Drohat, "Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 8091–8096, may 2012.
- [94] H. Youssoufian, H. H. Kazazian, D. G. Phillips, S. Aronis, G. Tsiftis, V. A. Brown, and S. E. Antonarakis, "Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots," *Nature*, vol. 324, pp. 380–382, nov 1986.
- [95] M. W. Simmen, "Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals," *Genomics*, vol. 92, pp. 33–40, jul 2008.

- [96] H. K. Long, D. Sims, A. Heger, N. P. Blackledge, C. Kutter, M. L. Wright, F. Grützner, D. T. Odom, R. Patient, C. P. Ponting, and R. J. Klose, “Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates,” eLife, vol. 2, pp. 1–19, feb 2013.
- [97] M. Gardiner-Garden and M. Frommer, “CpG Islands in vertebrate genomes,” Journal of Molecular Biology, vol. 196, pp. 261–282, jul 1987.
- [98] L. Han, B. Su, W.-H. Li, and Z. Zhao, “CpG island density and its correlations with genomic features in mammalian genomes,” Genome Biology, vol. 9, no. 5, p. R79, 2008.
- [99] A. P. Bird, M. Taggart, M. Frommer, O. J. Miller, and D. Macleod, “A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA,” Cell, vol. 40, pp. 91–99, jan 1985.
- [100] R. Stein, A. Razin, and H. Cedar, “In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells.,” Proceedings of the National Academy of Sciences, vol. 79, pp. 3418–3422, jun 1982.
- [101] I. Keshet, J. Yisraeli, and H. Cedar, “Effect of regional DNA methylation on gene expression.,” Proceedings of the National Academy of Sciences, vol. 82, pp. 2560–2564, may 1985.
- [102] J. Boyes and A. P. Bird, “Repression of genes by DNA methylation depends on CpG density and promoter strength : evidence for involvement of a methyl-CpG binding protein,” The EMBO journal, vol. 11, pp. 327–33, jan 1992.
- [103] C. L. Hsieh, “Dependence of transcriptional repression on CpG methylation density.,” Molecular and cellular biology, vol. 14, pp. 5487–94, aug 1994.
- [104] M. B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, and D. Schübeler, “DNA-binding factors shape the mouse methylome at distal regulatory regions,” Nature, vol. 480, pp. 490–495, dec 2011.
- [105] L. Han, I. G. Lin, and C.-L. Hsieh, “Protein Binding Protects Sites on Stable Episomes and in the Chromosome from De Novo Methylation,” Molecular and Cellular Biology, vol. 21, pp. 3416–3424, may 2001.
- [106] C. Kress, H. Thomassin, and T. Grange, “Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks,” Proceedings of the National Academy of Sciences, vol. 103, pp. 11112–11117, jul 2006.
- [107] Y. Liu, X. Zhang, R. M. Blumenthal, and X. Cheng, “A common mode of recognition for methylated CpG,” Trends in Biochemical Sciences, vol. 38, pp. 177–183, apr 2013.
- [108] A. M. Deaton and A. P. Bird, “CpG islands and the regulation of transcription,” Genes & Development, vol. 25, pp. 1010–1022, may 2011.
- [109] R. S. Illingworth, U. Gruenewald-Schneider, S. Webb, A. R. W. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird, “Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome,” PLoS Genetics, vol. 6, p. e1001134, sep 2010.
- [110] J. Autuoro, S. Pirnie, and G. Carmichael, “Long Noncoding RNAs in Imprinting and X Chromosome Inactivation,” Biomolecules, vol. 4, pp. 76–100, jan 2014.
- [111] A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D’Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. M. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello, “Conserved role of intragenic DNA methylation in regulating alternative promoters,” Nature, vol. 466, pp. 253–257, jul 2010.
- [112] J. Borgel, S. Guibert, Y. Li, H. Chiba, D. Schübeler, H. Sasaki, T. Forné, and M. Weber, “Targets and dynamics of promoter DNA methylation during early mouse development,” Nature Genetics, vol. 42, pp. 1093–1100, dec 2010.
- [113] P. A. Jones, “Functions of DNA methylation : islands, start sites, gene bodies and beyond,” Nature Reviews Genetics, vol. 13, pp. 484–492, may 2012.
- [114] C. M. Bender, M. L. Gonzalgo, F. A. Gonzales, C. T. Nguyen, K. D. Robertson, and P. A. Jones, “Roles of cell division and gene transcription in the methylation of CpG islands.,” Molecular and cellular biology, vol. 19, pp. 6690–6698, oct 1999.

- [115] C. E. Salem, I. D. C. Markl, C. M. Bender, F. A. Gonzales, P. A. Jones, and G. Liang, "PAX6 methylation and ectopic expression in human tumor cells," International journal of cancer, vol. 87, pp. 179–85, jul 2000.
- [116] M. Kulis, S. Heath, M. Bibikova, A. C. Queirós, A. Navarro, G. Clot, A. Martínez-Trillos, G. Castellano, I. Brun-Heath, M. Pinyol, S. Barberán-Soler, P. Papasaikas, P. Jares, S. Beà, D. Rico, S. Ecker, M. Rubio, R. Royo, V. Ho, B. Klotzle, L. Hernández, L. Conde, M. López-Guerra, D. Colomer, N. Villamor, M. Aymerich, M. Rozman, M. Bayes, M. Gut, J. L. Gelpí, M. Orozco, J.-B. Fan, V. Quesada, X. S. Puente, D. G. Pisano, A. Valencia, A. López-Guillermo, I. Gut, C. López-Otín, E. Campo, and J. I. Martín-Subero, "Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia," Nature Genetics, vol. 44, pp. 1236–1242, nov 2012.
- [117] K. E. Varley, J. Gertz, K. M. Bowling, S. L. Parker, T. E. Reddy, F. Pauli-Behn, M. K. Cross, B. A. Williams, J. A. Stamatoyannopoulos, G. E. Crawford, D. M. Absher, B. J. Wold, and R. M. Myers, "Dynamic DNA methylation across diverse human cell lines and tissues," Genome Research, vol. 23, pp. 555–567, mar 2013.
- [118] E. M. Wolff, H.-M. Byun, H. F. Han, S. Sharma, P. W. Nichols, K. D. Siegmund, A. S. Yang, P. A. Jones, and G. Liang, "Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer," PLoS Genetics, vol. 6, p. e1000917, apr 2010.
- [119] R. Tirado-Magallanes, K. Rebbani, R. Lim, S. Pradhan, and T. Benoukraf, "Whole genome DNA methylation : beyond genes silencing," Oncotarget, vol. 8, pp. 5629–5637, jan 2017.
- [120] S. Li, Y. Zhu, L. Zhi, X. Han, J. Shen, Y. Liu, J. Yao, and X. Yang, "DNA methylation variation trends during the embryonic development of chicken," PLoS ONE, vol. 11, no. 7, pp. 1–17, 2016.
- [121] G. Egger, G. Liang, A. Aparicio, and P. A. Jones, "Epigenetics in human disease and prospects for epigenetic therapy," Nature, vol. 429, pp. 457–463, may 2004.
- [122] B. T. Heijmans, E. W. Tobi, A. D. Stein, H. Putter, G. J. Blauw, E. S. Susser, P. E. Slagboom, and L. H. Lumey, "Persistent epigenetic differences associated with prenatal exposure to famine in humans.," Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 44, pp. 17046–9, 2008.
- [123] E. Heard and R. A. Martienssen, "Transgenerational Epigenetic Inheritance : Myths and Mechanisms," Cell, vol. 157, pp. 95–109, mar 2014.
- [124] L. Frésard, M. Morisson, J.-m. Brun, A. Collin, B. Pain, F. Minvielle, and F. Pitel, "Epigenetics and phenotypic variability : some interesting insights from birds," Genetics Selection Evolution, vol. 45, p. 16, jun 2013.
- [125] L. Lumey, A. D. Stein, H. S. Kahn, K. M. van der Pal-de Bruin, G. J. Blauw, P. A. Zybert, and E. S. Susser, "Cohort Profile : The Dutch Hunger Winter Families Study," International Journal of Epidemiology, vol. 36, pp. 1196–1204, dec 2007.
- [126] W. Reik, W. Dean, and J. Walter, "Epigenetic Reprogramming in Mammalian Development," Epigenetics, vol. 293, pp. 1089–1094, 2001.
- [127] S. Leroux, D. Gourichon, C. Leterrier, Y. Labrune, V. Coustham, S. Rivière, T. Zerjal, J. Coville, M. Morisson, F. Minvielle, and F. Pitel, "Embryonic environment and transgenerational effects in quail," Genetics Selection Evolution, vol. 49, p. 14, dec 2017.
- [128] C. Li, S. Guo, J. Gao, Y. Guo, E. Du, Z. Lv, and B. Zhang, "Maternal high-zinc diet attenuates intestinal inflammation by reducing DNA methylation and elevating H3K9 acetylation in the A20 promoter of offspring chicks," Journal of Nutritional Biochemistry, vol. 26, no. 2, pp. 173–183, 2014.
- [129] M. P. Lee, "Allele-specific gene expression and epigenetic modifications and their application to understanding inheritance and cancer," Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, vol. 1819, pp. 739–742, jul 2012.
- [130] K. Kerkel, A. Spadola, E. Yuan, J. Kosek, L. Jiang, E. Hod, K. Li, V. V. Murty, N. Schupf, E. Vilain, M. Morris, F. Haghighi, and B. Tycko, "Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation," Nature Genetics, vol. 40, pp. 904–908, jul 2008.

- [131] L. Frésard, S. Leroux, B. Servin, D. Gourichon, P. Dehais, M. San Cristobal, N. Marsaud, F. Vignoles, B. Bed'hom, J. L. Coville, F. Hormozdiari, C. Beaumont, T. Zerjal, A. Vignal, M. Morisson, S. Lagarrigue, and F. Pitel, "Transcriptome-wide investigation of genomic imprinting in chicken," *Nucleic Acids Research*, vol. 42, no. 6, pp. 3768–3782, 2014.
- [132] Q. Wang, K. Li, D. Zhang, J. Li, G. Xu, J. Zheng, N. Yang, and L. Qu, "Next-Generation Sequencing Techniques Reveal that Genomic Imprinting Is Absent in Day-Old Gallus gallus domesticus Brains," *PLOS ONE*, vol. 10, p. e0132345, jul 2015.
- [133] Z. Zhuo, S. J. Lamont, and B. Abasht, "RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken," *Scientific Reports*, vol. 7, p. 11944, dec 2017.
- [134] R. Shoemaker, J. Deng, W. Wang, and K. Zhang, "Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome," *Genome Research*, vol. 20, pp. 883–889, jul 2010.
- [135] W. Xie, C. L. Barr, A. Kim, F. Yue, A. Y. Lee, J. Eubanks, E. L. Dempster, and B. Ren, "Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome," *Cell*, vol. 148, pp. 816–831, feb 2012.
- [136] X. Wang, J. H. Werren, and A. G. Clark, "Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in *Nasonia*," *PLoS Biology*, vol. 14, p. e1002500, jul 2016.
- [137] B. Tycko, "Allele-specific DNA methylation : beyond imprinting," *Human Molecular Genetics*, vol. 19, pp. R210–R220, oct 2010.
- [138] M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, and J. R. Ecker, "Human body epigenome maps reveal noncanonical DNA methylation variation," *Nature*, vol. 523, pp. 212–216, jul 2015.
- [139] J. N. Hutchinson, T. Raj, J. Fagerness, E. Stahl, F. T. Vioria, A. Gimelbrant, J. Seddon, M. Daly, A. Chess, and R. Plenge, "Allele-Specific Methylation Occurs at Genetic Variants Associated with Complex Disease," *PLoS ONE*, vol. 9, p. e98464, jun 2014.
- [140] D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, and J. C. Vamosi, "Sex Determination : Why So Many Ways of Doing It ?," *PLoS Biology*, vol. 12, p. e1001899, jul 2014.
- [141] D. E. Irwin, "Sex chromosomes and speciation in birds and other ZW systems," *Molecular Ecology*, pp. 0–2, feb 2018.
- [142] Y. Xiong, X. Chen, Z. Chen, X. Wang, S. Shi, X. Wang, J. Zhang, and X. He, "RNA sequencing shows no dosage compensation of the active X-chromosome," *Nature Genetics*, vol. 42, pp. 1043–1047, dec 2010.
- [143] P. Julien, D. Brawand, M. Soumillon, A. Necsulea, A. Liechti, F. Schütz, T. Daish, F. Grützner, and H. Kaessmann, "Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage Compensation," *PLoS Biology*, vol. 10, p. e1001328, may 2012.
- [144] H. Ellegren, L. Hultin-Rosenberg, B. Brunström, L. Dencker, K. Kultima, and B. Scholz, "Faced with inequality : chicken do not have a general dosage compensation of sex-linked genes," *BMC Biology*, vol. 5, no. 1, p. 40, 2007.
- [145] H. Ellegren, "The Evolutionary Genomics of Birds," *Annual Review of Ecology, Evolution, and Systematics*, vol. 44, pp. 239–259, nov 2013.
- [146] F. Zimmer, P. W. Harrison, C. Dessimoz, and J. E. Mank, "Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome," *Genome Biology and Evolution*, vol. 8, pp. 1233–1242, apr 2016.
- [147] E. Melamed and A. P. Arnold, "Regional differences in dosage compensation on the chicken Z chromosome," *Genome Biology*, vol. 8, no. 9, p. R202, 2007.
- [148] Y. Itoh, K. Replogle, Y. H. Kim, J. Wade, D. F. Clayton, and A. P. Arnold, "Sex bias and dosage compensation in the zebra finch versus chicken genomes : General and specialized patterns among birds," *Genome Research*, vol. 20, pp. 512–518, apr 2010.
- [149] L. Bisoni, L. Batlle-Morera, A. P. Bird, M. Suzuki, and H. A. McQueen, "Female-specific hyperacetylation of histone H4 in the chicken Z chromosome," *Chromosome Research*, vol. 13, pp. 205–214, feb 2005.

- [150] E. C. Small, L. Xi, J.-P. Wang, J. Widom, and J. D. Licht, “Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity,” Proceedings of the National Academy of Sciences, vol. 111, pp. E2462–E2471, jun 2014.
- [151] International Chicken Genome Sequencing Consortium, “Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.,” Nature, vol. 432, pp. 695–716, dec 2004.
- [152] F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors.,” Proceedings of the National Academy of Sciences of the United States of America, vol. 74, pp. 5463–7, dec 1977.
- [153] W. C. Warren, L. W. Hillier, C. Tomlinson, P. Minx, M. Kremitzki, T. Graves, C. Markovic, N. Bouk, K. D. Pruitt, F. Thibaud-Nissen, V. Schneider, T. A. Mansour, C. T. Brown, A. Zimin, R. Hawken, M. Abrahamsen, A. B. Pyrkosz, M. Morisson, V. Fillon, A. Vignal, W. Chow, K. Howe, J. E. Fulton, M. M. Miller, P. Lovell, C. V. Mello, M. Wirthlin, A. S. Mason, R. Kuo, D. W. Burt, J. B. Dodgson, and H. H. Cheng, “A New Chicken Genome Assembly Provides Insight into Avian Genome Structure.,” G3 (Bethesda, Md.), vol. 7, pp. 109–117, jan 2017.
- [154] A. Rhoads and K. F. Au, “PacBio Sequencing and Its Applications,” Genomics, Proteomics and Bioinformatics, vol. 13, no. 5, pp. 278–289, 2015.
- [155] F. Krueger, B. Kreck, A. Franke, and S. R. Andrews, “DNA methylome analysis using short bisulfite sequencing data,” Nature Methods, vol. 9, pp. 145–151, jan 2012.
- [156] B. Y. Michel, M. Cuendet, N. P. Barthes, A. Burger, P. Bertrand, and N. Martinet, “Détection des modifications de la méthylation de l’ADN et de la chromatine,” Revue Francophone des Laboratoires, vol. 2015, pp. 55–62, jun 2015.
- [157] M. A. Urich, J. R. Nery, R. Lister, R. J. Schmitz, and J. R. Ecker, “MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing,” Nature Protocols, vol. 10, no. 3, pp. 475–483, 2015.
- [158] S.-A. David, M. Mersch, S. Foissac, A. Collin, F. Pitel, and V. Coustham, “Genome-Wide Epigenetic Studies in Chicken : A Review,” Epigenomes, vol. 1, p. 20, nov 2017.
- [159] A. Meissner, “Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis,” Nucleic Acids Research, vol. 33, pp. 5868–5877, oct 2005.
- [160] H. Gu, Z. D. Smith, C. Bock, P. Boyle, A. Gnirke, and A. Meissner, “Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling,” Nature Protocols, vol. 6, pp. 468–481, mar 2011.
- [161] F. Krueger and S. R. Andrews, “Bismark : a flexible aligner and methylation caller for Bisulfite-Seq applications,” Bioinformatics, vol. 27, pp. 1571–1572, jun 2011.
- [162] K. Wreczycka, A. Goidschan, D. Yusuf, B. Grüning, Y. Assenov, and A. Akalin, “Strategies for analyzing bisulfite sequencing data,” Journal of Biotechnology, vol. 261, pp. 105–115, nov 2017.
- [163] M. Martin, “Cutadapt removes adapter sequences from high-throughput sequencing reads,” EMBnet.journal, vol. 17, p. 10, may 2011.
- [164] S. R. Andrews, “FastQC : a quality control tool for high throughput sequence data,” 2010.
- [165] Z. Sun, J. Cunningham, S. Slager, and J.-P. Kocher, “Base resolution methylome profiling : considerations in platform selection, data preprocessing and analysis.,” Epigenomes, vol. 7, no. 5, pp. 813–28, 2015.
- [166] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, “Ultrafast and memory-efficient alignment of short DNA sequences to the human genome,” Genome Biology, vol. 10, no. 3, p. R25, 2009.
- [167] M. Burrows and D. J. Wheeler, “A Block Sorting Lossless Data Compression Algorithm,” tech. rep., Digital Equipment Corporation, Palo Alto, CA, 1994.
- [168] X. Peng, J. Wang, Z. Zhang, Q. Xiao, M. Li, and Y. Pan, “Re-alignment of the unmapped reads with base quality score,” BMC Bioinformatics, vol. 16, no. Suppl 5, p. S8, 2015.
- [169] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, “The Sequence Alignment/Map format and SAMtools,” Bioinformatics, vol. 25, pp. 2078–2079, aug 2009.

- [170] A. Akalin, M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, and C. E. Mason, “methylKit : a comprehensive R package for the analysis of genome-wide DNA methylation profiles,” Genome Biology, vol. 13, no. 10, p. R87, 2012.
- [171] A. Breiling and F. Lyko, “Epigenetic regulatory functions of DNA modifications : 5-methylcytosine and beyond,” Epigenetics & Chromatin, vol. 8, no. 1, p. 24, 2015.
- [172] D. P. Genereux, W. C. Johnson, A. F. Burden, R. Stoger, and C. D. Laird, “Errors in the bisulfite conversion of DNA : modulating inappropriate- and failed-conversion frequencies,” Nucleic Acids Research, vol. 37, pp. 5235–5235, aug 2009.
- [173] H. Feng, K. N. Conneely, and H. Wu, “A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data,” Nucleic Acids Research, vol. 42, pp. 1–11, apr 2014.
- [174] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. C. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker, “Human DNA methylomes at base resolution show widespread epigenomic differences.,” Nature, vol. 462, pp. 315–22, nov 2009.
- [175] E. R. Holzinger and M. D. Ritchie, “Integrating heterogeneous high-throughput data for meta-dimensional pharmacogenomics and disease-related studies,” Pharmacogenomics, vol. 13, pp. 213–222, jan 2012.
- [176] E. R. Holzinger, S. M. Dudek, A. T. Frase, R. M. Krauss, M. W. Medina, and M. D. Ritchie, “ATHENA : a tool for meta-dimensional analysis applied to genotypes and gene expression data to predict HDL cholesterol levels.,” Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, vol. 29, pp. 385–96, nov 2013.
- [177] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, “Methods of integrating data to uncover genotype-phenotype interactions,” Nature Reviews Genetics, vol. 16, no. 2, pp. 85–97, 2015.
- [178] N. Bumstead and P. A. Barrow, “Genetics of Resistance to Salmonella Typhimurium in Newly Hatched Chicks,” British Poultry Science, vol. 29, no. 3, pp. 521–529, 1988.
- [179] L. Frésard, S. Leroux, P.-F. Roux, C. Klopp, S. Fabre, D. Esquerré, P. Dehais, A. Djari, D. Gourichon, S. Lagarrigue, and F. Pitel, “Genome-Wide Characterization of RNA Editing in Chicken Embryos Reveals Common Features among Vertebrates,” PLoS ONE, vol. 10, p. e0126776, may 2015.
- [180] N. Campbell, Biologie. 3e édition ed., 1995.
- [181] S. M. Smith and W. W. Vale, “The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress,” Dialogues in Clinical Neuroscience, vol. 8, no. 4, pp. 383–395, 2006.
- [182] S. F. Jastrebski, S. J. Lamont, and C. J. Schmidt, “Chicken hepatic response to chronic heat stress using integrated transcriptome and metabolome analysis,” PLoS ONE, vol. 12, no. 7, pp. 1–15, 2017.
- [183] A. Bordas, M. Tixier-Boichard, and P. Merat, “Direct And Correlated Responses To Divergent Selection For Residual Food Intake In Rhode Island Red Laying Hens,” British Poultry Science, vol. 33, no. 4, pp. 741–754, 1992.
- [184] L. Pei, J.-H. Choi, J. Liu, E.-J. Lee, B. McCarthy, J. M. Wilson, E. Speir, F. Awan, H. Tae, G. Arthur, J. L. Schnabel, K. H. Taylor, X. Wang, D. Xu, H.-F. Ding, D. H. Munn, C. Caldwell, and H. Shi, “Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia,” Epigenetics, vol. 7, pp. 567–578, jun 2012.
- [185] X.-L. Yuan, N. Gao, Y. Xing, H.-B. Zhang, A.-L. Zhang, J. Liu, J.-L. He, Y. Xu, W.-M. Lin, Z.-M. Chen, H. Zhang, Z. Zhang, and J.-Q. Li, “Profiling the genome-wide DNA methylation pattern of porcine ovaries using reduced representation bisulfite sequencing,” Scientific Reports, vol. 6, p. 22138, apr 2016.
- [186] X. Zhang, S. Zhang, L. Ma, E. Jiang, H. Xu, R. Chen, Q. Yang, H. Chen, Z. Li, and X. Lan, “Reduced representation bisulfite sequencing (RRBS) of dairy goat mammary glands reveals DNA methylation profiles of integrated genome-wide and critical milk-related genes.,” Oncotarget, vol. 8, no. 70, pp. 115326–115344, 2017.
- [187] A. R. Quinlan and I. M. Hall, “BEDTools : a flexible suite of utilities for comparing genomic features,” Bioinformatics, vol. 26, pp. 841–842, mar 2010.

- [188] J. Mariette, F. Escudié, P. Bardou, I. Nabihoudine, C. Noirot, M.-S. Trotard, C. Gaspin, and C. Klopp, “Jflow : a workflow management system for web applications,” *Bioinformatics*, vol. 32, pp. 456–458, feb 2016.
- [189] D. Li, B. Zhang, X. Xing, and T. Wang, “Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation,” *Methods*, vol. 72, pp. 29–40, jan 2015.
- [190] M. Zhang, F.-B. Yan, F. Li, K.-R. Jiang, D.-H. Li, R.-L. Han, Z.-J. Li, R.-R. Jiang, X.-J. Liu, X.-T. Kang, and G.-R. Sun, “Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens,” *Scientific Reports*, vol. 7, p. 45564, apr 2017.
- [191] Z. Zhang, H. Du, L. Bai, C. Yang, Q. Li, X. Li, M. Qiu, C. Yu, Z. Jiang, X. Jiang, L. Liu, C. Hu, B. Xia, X. Xiong, X. Song, and X. Jiang, “Whole genome bisulfite sequencing reveals unique adaptations to high-altitude environments in Tibetan chickens,” *PLOS ONE*, vol. 13, p. e0193597, mar 2018.
- [192] M. J. Ziller, K. D. Hansen, A. Meissner, and M. J. Aryee, “Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing,” *Nature Methods*, vol. 12, pp. 230–232, nov 2015.
- [193] Q. Li, N. Li, X. Hu, J. Li, Z. Du, L. Chen, G. Yin, J. Duan, H. Zhang, Y. Zhao, J. Wang, and N. Li, “Genome-Wide Mapping of DNA Methylation in Chicken,” *PLoS ONE*, vol. 6, p. e19428, may 2011.
- [194] Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate : a Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society Seribes B (Methodological)*, vol. 135, no. 1, pp. 289–300, 1995.
- [195] M. D. Robinson, A. Kahraman, C. W. Law, H. Lindsay, M. Nowicka, L. M. Weber, and X. Zhou, “Statistical methods for detecting differentially methylated loci and regions,” *Frontiers in Genetics*, vol. 5, pp. 1–7, sep 2014.
- [196] J. F. Degner, J. C. Marioni, A. A. Pai, J. K. Pickrell, E. Nkadori, Y. Gilad, and J. K. Pritchard, “Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data,” *Bioinformatics*, vol. 25, pp. 3207–3212, dec 2009.
- [197] S. E. Castel, A. Levy-Moonshine, P. Mohammadi, E. Banks, and T. Lappalainen, “Tools and best practices for data processing in allelic expression analysis,” *Genome Biology*, vol. 16, p. 195, dec 2015.
- [198] W. A. Cheung, X. Shao, A. Morin, V. Siroux, T. Kwan, B. Ge, D. Aïssi, L. Chen, L. Vasquez, F. Allum, F. Guénard, E. Bouzigon, M.-M. Simon, E. Boulrier, A. Redensek, S. Watt, A. Datta, L. Clarke, P. Flicek, D. Mead, D. S. Paul, S. Beck, G. Bourque, M. Lathrop, A. Tchernof, M.-C. Vohl, F. Demenais, I. Pin, K. Downes, H. G. Stunnenberg, N. Soranzo, T. Pastinen, and E. Grundberg, “Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome.,” *Genome biology*, vol. 18, p. 50, mar 2017.
- [199] F. Krueger and S. R. Andrews, “SNPsplit : Allele-specific splitting of alignments between genomes with known SNP genotypes,” *F1000Research*, vol. 5, p. 1479, jul 2016.
- [200] P. Wulfridge, B. Langmead, A. P. Feinberg, and K. Hansen, “Choice of reference genome can introduce massive bias in bisulfite sequencing data,” *bioRxiv*, pp. 1–31, 2016.
- [201] M. El-Kazzi, A. Bordas, G. Gandemer, and F. Minvielle, “Divergent selection for residual food intake in Rhode Island red egg-laying lines : gross carcass composition, carcass adiposity and lipid contents of tissues.,” *British poultry science*, vol. 36, pp. 719–28, dec 1995.
- [202] K. P. Leung, P. D’Arpa, A. K. Seth, M. R. Geringer, M. Jett, W. Xu, S. J. Hong, R. D. Galiano, T. Chen, and T. A. Mustoe, “Dermal wound transcriptomic responses to Infection with *Pseudomonas aeruginosa* versus *Klebsiella pneumoniae* in a rabbit ear wound model,” *BMC Clinical Pathology*, vol. 14, p. 20, dec 2014.
- [203] S. Vuillaumier-Barrot, C. Bouchet-Séraphin, M. Chelbi, L. Devisme, S. Quentin, S. Gazal, A. Laquerrière, C. Fallet-Bianco, P. Loget, S. Odent, D. Carles, A. Bazin, J. Aziza, A. Clemenson, F. Guimiot, M. Bonnière, S. Monnot, C. Bole-Feysot, J.-P. Bernard, L. Loeuillet, M. Gonzales, K. Socha, B. Grandchamp, T. Attié-Bitach, F. Encha-Razavi, and N. Seta, “Identification of Mutations in TMEM5 and ISPD as a Cause of Severe Cobblestone Lissencephaly,” *The American Journal of Human Genetics*, vol. 91, pp. 1135–1143, dec 2012.

- [204] Z. Lonsdale, K. Lee, M. Kiriakidu, H. Amarasinghe, D. Nathanael, C. J. O'Connor, and E. B. Mallon, "Allele specific expression and methylation in the bumblebee, *Bombus terrestris*," PeerJ, vol. 5, p. e3798, sep 2017.
- [205] A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander, "Genome-scale DNA methylation maps of pluripotent and differentiated cells," Nature, vol. 454, pp. 766–770, aug 2008.
- [206] T. Hartung, L. Zhang, R. Kanwar, I. Khrebtukova, M. Reinhardt, C. Wang, T. M. Therneau, M. S. Banck, G. P. Schroth, and A. S. Beutler, "Diametrically opposite methylome-transcriptome relationships in high- and low-CpG promoter genes in postmitotic neural rat tissue," Epigenetics, vol. 7, pp. 421–428, may 2012.
- [207] A. Chatterjee, Y. Ozaki, P. A. Stockwell, J. A. Horsfield, I. M. Morison, and S. Nakagawa, "Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing," Epigenetics, vol. 8, pp. 979–989, sep 2013.
- [208] R. Doherty and C. Couldrey, "Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock : a technical assessment," Frontiers in genetics, vol. 5, p. 126, may 2014.
- [209] J. Su, Y. Wang, X. Xing, J. Liu, and Y. Zhang, "Genome-wide analysis of DNA methylation in bovine placentas," BMC Genomics, vol. 15, no. 1, p. 12, 2014.
- [210] J.-R. Lee, C. Hong, J.-W. Moon, Y.-D. Jung, D.-S. Kim, T.-H. Kim, J.-A. Gim, J.-H. Bae, Y. Choi, J. Eo, Y.-J. Kwon, S. Song, J. Ko, Y. Yang, H.-K. Lee, K.-D. Park, K. Ahn, K.-T. Do, H.-S. Ha, K. Han, J. Yi, H.-J. Cha, B.-W. Cho, J. Bhak, and H.-S. Kim, "Genome-wide analysis of DNA methylation patterns in horse," BMC Genomics, vol. 15, no. 1, p. 598, 2014.
- [211] Y. Hao, Y. Cui, and X. Gu, "Genome-wide DNA methylation profiles changes associated with constant heat stress in pigs as measured by bisulfite sequencing," Scientific Reports, vol. 6, p. 27507, jul 2016.
- [212] Z. D. Smith, H. Gu, C. Bock, A. Gnirke, and A. Meissner, "High-throughput bisulfite sequencing in mammalian genomes," Methods, vol. 48, pp. 226–232, jul 2009.
- [213] Y. Hu, H. Xu, Z. Li, X. Zheng, X. Jia, Q. Nie, and X. Zhang, "Comparison of the Genome-Wide DNA Methylation Profiles between Fast-Growing and Slow-Growing Broilers," PLoS ONE, vol. 8, p. e56411, feb 2013.
- [214] R. Edgar, P. P. Tan, E. Portales-Casamar, and P. Pavlidis, "Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression," Epigenetics & Chromatin, vol. 7, no. 1, p. 28, 2014.
- [215] I. P. Ioshikhes and M. Q. Zhang, "Large-scale human promoter mapping using CpG islands," Nature Genetics, vol. 26, pp. 61–63, sep 2000.
- [216] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, "STAR : ultrafast universal RNA-seq aligner," Bioinformatics, vol. 29, pp. 15–21, jan 2013.
- [217] B. Li and C. N. Dewey, "RSEM : accurate transcript quantification from RNA-Seq data with or without a reference genome," BMC Bioinformatics, vol. 12, no. 1, p. 323, 2011.
- [218] Y.-Z. Huang, J.-J. Sun, L.-Z. Zhang, C.-J. Li, J. E. Womack, Z.-J. Li, X.-Y. Lan, C.-Z. Lei, C.-L. Zhang, X. Zhao, and H. Chen, "Genome-wide DNA Methylation Profiles and Their Relationships with mRNA and the microRNA Transcriptome in Bovine Muscle Tissue (*Bos taurine*)," Scientific Reports, vol. 4, p. 6546, may 2015.
- [219] M. P. Lisanti, K. Reeves, M. Peiris-Pagès, A. L. Chadwick, R. Sanchez-Alvarez, A. Howell, U. E. Martinez-Outschoorn, and F. Sotgia, "JNK1 stress signaling is hyper-activated in high breast density and the tumor stroma : Connecting fibrosis, inflammation, and stemness for cancer prevention," Cell Cycle, vol. 13, pp. 580–599, feb 2014.
- [220] C. Fasquelle, A. Sartelet, W. Li, M. Dive, N. Tamma, C. Michaux, T. Druet, I. J. Huijbers, C. M. Isacke, W. Coppieters, M. Georges, and C. Charlier, "Balancing Selection of a Frame-Shift Mutation in the MRC2 Gene Accounts for the Outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle," PLoS Genetics, vol. 5, p. e1000666, sep 2009.
- [221] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR : a Bioconductor package for differential expression analysis of digital gene expression data," Bioinformatics, vol. 26, pp. 139–140, jan 2010.

- [222] B. Du, W. P. Cawthorn, A. Su, C. R. Doucette, Y. Yao, N. Hemati, S. Kampert, C. McCoin, D. T. Broome, C. J. Rosen, G. Yang, and O. A. MacDougald, "The Transcription Factor Paired-Related Homeobox 1 (Prrx1) Inhibits Adipogenesis by Activating Transforming Growth Factor- β (TGF β) Signaling," Journal of Biological Chemistry, vol. 288, pp. 3036–3047, feb 2013.
- [223] R. R. Nepomuceno and a. J. Tenner, "C1qRP, the C1q receptor that enhances phagocytosis, is detected specifically in human cells of myeloid lineage, endothelial cells, and platelets.," Journal of immunology (Baltimore, Md. : 1950), vol. 160, no. 4, pp. 1929–1935, 1998.
- [224] M. C. Greenlee-Wacker, M. D. Galvan, and S. S. Bohlson, "CD93 : Recent Advances and Implications in Disease," Current Drug Targets, vol. 13, pp. 411–420, mar 2012.
- [225] V. Calvanese, A. F. Fernández, R. G. Urduingio, B. Suárez-Alvarez, C. Mangas, V. Pérez-García, C. Bueno, R. Montes, V. Ramos-Mejía, P. Martínez-Cambor, C. Ferrero, Y. Assenov, C. Bock, P. Menendez, A. C. Carrera, C. Lopez-Larrea, and M. F. Fraga, "A promoter DNA demethylation landscape of human hematopoietic differentiation," Nucleic Acids Research, vol. 40, pp. 116–131, jan 2012.
- [226] T. Li, Y. Shi, P. Wang, L. M. Guachalla, B. Sun, T. Joerss, Y.-S. Chen, M. Groth, A. Krueger, M. Platzer, Y.-G. Yang, K. L. Rudolph, and Z.-Q. Wang, "Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay," The EMBO Journal, vol. 34, pp. 1630–1647, jun 2015.
- [227] D. Wirth and P. Gustin, "Les protéines de choc thermique (heat shock proteins)," Ann. Méd. Vét., vol. 06, no. 146, pp. 201–216, 2002.
- [228] W. E. Ek, E. W. Tobi, M. Ahsan, E. Lampa, E. Ponzi, S. A. Kyrtopoulos, P. Georgiadis, L. Lumey, B. T. Heijmans, M. Botsivali, I. A. Bergdahl, T. Karlsson, M. Rask-Andersen, D. Palli, E. Ingelsson, Å. K. Hedman, L. M. Nilsson, P. Vineis, L. Lind, J. M. Flanagan, and Å. Johansson, "Tea and coffee consumption in relation to DNA methylation in four European cohorts," Human Molecular Genetics, vol. 26, pp. 3221–3231, aug 2017.
- [229] E.-J. Seo, C.-F. Wu, Z. Ali, Y.-H. Wang, S. I. Khan, L. A. Walker, I. A. Khan, and T. Efferth, "Both Phenolic and Non-phenolic Green Tea Fractions Inhibit Migration of Cancer Cells," Frontiers in Pharmacology, vol. 7, pp. 1–16, nov 2016.
- [230] D. Li, R. Zhang, W. Zhu, Y. Xue, Y. Zhang, Q. Huang, M. Liu, and Y. Liu, "S100A16 inhibits osteogenesis but stimulates adipogenesis," Molecular Biology Reports, vol. 40, pp. 3465–3473, may 2013.
- [231] D. Chen, L. Luo, and C. Liang, "Aberrant S100A16 expression might be an independent prognostic indicator of unfavorable survival in non-small cell lung adenocarcinoma," PLOS ONE, vol. 13, p. e0197402, may 2018.
- [232] M. S. Almén, J. A. Jacobsson, G. Moschonis, C. Benedict, G. P. Chrousos, R. Fredriksson, and H. B. Schiöth, "Genome wide analysis reveals association of a FTO gene variant with epigenetic changes," Genomics, vol. 99, pp. 132–137, mar 2012.
- [233] A. E. Gururaj, R. R. Singh, S. K. Rayala, C. Holm, P. den Hollander, H. Zhang, S. Balasenthil, A. H. Talukder, G. Landberg, and R. Kumar, "MTA1, a transcriptional activator of breast cancer amplified sequence 3," Proceedings of the National Academy of Sciences, vol. 103, pp. 6670–6675, apr 2006.
- [234] M. Codrich, M. Bertuzzi, R. Russo, M. Francescato, S. Espinoza, L. Zentilin, M. Giacca, D. Cesselli, A. P. Beltrami, P. Ascenzi, S. Zucchelli, F. Persichetti, G. Leanza, and S. Gustincich, "Neuronal hemoglobin affects dopaminergic cells' response to stress," Cell Death & Disease, vol. 8, pp. e2538–e2538, jan 2017.
- [235] P. May and E. May, "Rôles de p53 dans le cycle cellulaire, l'apoptose et la cancer," Revue Française des Laboratoires n°311, pp. 35–42, 1999.
- [236] M. Affolter, J. Côté, J. Renaud, and A. Ruiz-Carrillo, "Regulation of histone and beta A-globin gene expression during differentiation of chicken erythroid cells.," Molecular and Cellular Biology, vol. 7, pp. 3663–3672, oct 1987.
- [237] S. Jahan, W. Xu, S. He, C. Gonzalez, G. P. Delcuve, and J. R. Davie, "The chicken erythrocyte epigenome," Epigenetics & Chromatin, vol. 9, p. 19, dec 2016.
- [238] S. Kurdyukov and M. Bullock, "DNA Methylation Analysis : Choosing the Right Method," Biology, vol. 5, p. 3, jan 2016.

- [239] X. Shao, C. Zhang, M.-A. Sun, X. Lu, and H. Xie, “Deciphering the heterogeneity in DNA methylation patterns during stem cell differentiation and reprogramming,” *BMC Genomics*, vol. 15, no. 1, p. 978, 2014.
- [240] S. Schbath, V. Martin, M. Zytnicki, J. Fayolle, V. Loux, and J.-F. Gibrat, “Mapping Reads on a Genomic Sequence : An Algorithmic Overview and a Practical Comparative Analysis,” *Journal of Computational Biology*, vol. 19, pp. 796–813, jun 2012.
- [241] A. Merkel, M. Fernandez-Callejo, E. Casals, S. Marco-Sola, R. Schuyler, I. G. Gut, and S. C. Heath, “GEMBS - high through-put processing for DNA methylation data from Whole Genome Bisulfite Sequencing (WGBS),” *bioRxiv*, p. 201988, 2017.
- [242] S. Marco-Sola, M. Sammeth, R. Guigó, and P. Ribeca, “The GEM mapper : fast, accurate and versatile alignment by filtration,” *Nature Methods*, vol. 9, p. 1185, oct 2012.
- [243] Y. Liu, K. D. Siegmund, P. W. Laird, and B. P. Berman, “Bis-SNP : Combined DNA methylation and SNP calling for Bisulfite-seq data,” *Genome Biology*, vol. 13, no. 7, p. R61, 2012.
- [244] S. Lou, H.-M. Lee, H. Qin, J.-W. Li, Z. Gao, X. Liu, L. L. Chan, V. KL Lam, W.-Y. So, Y. Wang, S. Lok, J. Wang, R. C. Ma, S. K.-W. Tsui, J. C. Chan, T.-F. Chan, and K. Y. Yip, “Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation,” *Genome Biology*, vol. 15, p. 408, jul 2014.
- [245] D. Anastasiadi, A. Esteve-Codina, and F. Piferrer, “Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species,” *Epigenetics & Chromatin*, vol. 11, p. 37, dec 2018.
- [246] Y. Zhang, R. Wilson, J. Heiss, L. P. Breitling, K. U. Saum, B. Schöttker, B. Holleczeck, M. Waldenberger, A. Peters, and H. Brenner, “DNA methylation signatures in peripheral blood strongly predict all-cause mortality,” *Nature Communications*, vol. 8, pp. 1–11, 2017.
- [247] Z.-Y. Zhou, Y. Hu, A. Li, Y.-J. Li, H. Zhao, S.-Q. Wang, N. O. Otecko, D. Zhang, J.-H. Wang, Y. Liu, D. M. Irwin, Y. Qin, and Y.-P. Zhang, “Genome wide analyses uncover allele-specific RNA editing in human and mouse,” *Nucleic Acids Research*, no. August, pp. 1–10, 2018.
- [248] T. K. Kelly, Y. Liu, F. D. Lay, G. Liang, B. P. Berman, and P. a. Jones, “Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules,” *Genome Research*, vol. 22, pp. 2497–2506, dec 2012.
- [249] J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, and W. Timp, “Detecting DNA cytosine methylation using nanopore sequencing,” *Nature Methods*, vol. 14, no. 4, pp. 407–410, 2017.
- [250] S.-A. David, B. Piégu, C. Hennequet-Antier, M. Pannetier, T. Aguirre-Lavin, S. Crochet, T. Bordeau, N. Couroussé, A. Brionne, Y. Bigot, A. Collin, and V. Coustham, “An Assessment of Fixed and Native Chromatin Preparation Methods to Study Histone Post-Translational Modifications at a Whole Genome Scale in Skeletal Muscle Tissue,” *Biological Procedures Online*, vol. 19, p. 10, dec 2017.

